DOCUMENT RESUME

ED 043 651 24 TE 499 848

AUTHOR	Kyme, George H.
TITLE	The Value of Aesthetic Judgments in Music in the Assessment of Musicality of Elementary School Children. Final Report.
INSTITUTION	California Univ., Berkeley.
SPONS AGENCY	Office of Education (DHEW), Washington, D.C. Bureau of Research.
BUREAU NO	BR-8-I-146
PUB DATE	Jul 70
GRANT	OEG-9-9-140146-0023(057)
NOTE	83p.
EDRS PRICE	EDRS Price MF-$0.50 HC-$4.25
DESCRIPTORS	Cultural Enrichment, *Elementary Education, *Listening Comprehension, Music, *Music Appreciation, *Music Education, Perception, *Testing

ABSTRACT

A test of musical achievement, the central criterion of which was musical sensitivity as evidenced by the ability to make aesthetic judgments, was empirically validated and field-tested upon elementary school children in the Manhattanville Music Curriculum Program as well as in the California Music Educators Statewide Music Testing Program. The test was comprised of excerpts of compositions by recognized composers and a mutilation of each. Reactions of the Manhattanville children, who received training in creating and experiencing contemporary music, differed from those of their California peers who received other types of musical training. Some of the observations made were that (1) musical children of the same age score similiarly though drawn from divergent programs; (2) Manhattanville subjects were more perceptive than their California counterparts but less likely to agree that the original composition excels its mutilation; and (3) individual differences are greater within samples than between them. (Tables, test items and criteria, and a brief bibliography are provided.) (Author/MF)

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
OFFICE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL OFFICE OF EDUCATION POSITION OR POLICY.

ED043651

BR 8-I-146
PA 24

THE VALUE OF AESTHETIC JUDGMENTS IN MUSIC IN THE ASSESSMENT OF MUSICALITY OF ELEMENTARY SCHOOL CHILDREN

Project No. 8-1-146
Grant No. OEG-9-9-140146-0023(057)

George H. Kyme

July, 1970

The research reported herein was performed pursuant to a contract with the Office of Education, U. S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

TE 499 848

University of California
Berkeley, California

ABSTRACT

THE VALUE OF AESTHETIC JUDGMENTS IN MUSIC IN THE ASSESSMENT OF MUSICALITY OF ELEMENTARY SCHOOL CHILDREN

George H. Kyme
The University of California, Berkeley

1. Abstract.

(a) Objectives: It was the purpose of this research to develop a test of musical achievement whose central criterion is musical sensitivity as evidenced by the ability to make aesthetic judgments in music. The test was empirically validated and field-tested upon elementary school children in the Manhattanville Music Curriculum program as well as in the California Music Educators Statewide Music Testing Program.

Rationale: "The purpose of music education throughout the grades," members of the Yale Seminar on Music Education agreed, "is to develop musicality." "Musicality may be defined as the ability to express through time and pitch the mental image of a musical idea. Conversely, it is the ability to grasp in its entirety a musical idea which has been expressed." This latter definition is not in opposition to the purpose of music education as expressed by Broudy (12) in the Tanglewood Symposium to the point that "Music education is, in the final analysis, aesthetic education." The word "aesthetic" in its Greek origin, aisthetes, means "one who perceives.

Contribution to Education: Because classroom courses in music whose purpose is to make music education aesthetic education have been the exception, testing in the domain of aesthetic sensitivity has not progressed at the rate it has with other fields. There are few standardized tests for achievement, and these have not won any widespread acceptance. Consequently, music educators and researchers who would institute or evaluate innovative programs are hard put to determine the effectiveness of such programs either in terms of stated objectives or in relationship to other avenues of music education prevalent in the schools. It is the intent of this endeavor to develop a much needed test of musical achievement by exploring the aesthetic responses of children in the elementary schools, to determine what the typical aesthetic responses are and what the normal rate of development in aesthetic sensitivity to music is, and, finally, to discover in which ways those students rated by their teachers as being most musical will differ in musical sensitivity from those classmates assessed as being least musical.

(b) Procedure: A test of aesthetic judgments in music was developed which, in the Hevner tradition, was comprised of paired short pieces, one of which was a systematic mutilation of its essence (i.e., melody, harmony, rhythm, form, tone-color, or musical style.) The test was administered to 1800 children in the Madera, California, PACE project which was currently evaluating the Orff and Kodaly techniques for developing musicality, as well as to 1600 children participating in the San Diego County Title III project Developing Impetus for the Creative Arts. This project proposed the development of musicality through conservatory training and the utilization of demonstration performances by professionals in the community. The test items were empirically validated in that only those items which differentiated significantly between students rated as being most musical and those rated as being least musical were selected for the final test. The revised test was used to evaluate growth in musical sensitivity of students in the Manhattanville Music Curriculum Program. In doing so, the test was given to 800 students in that program to discover how children receiving training in creating and experiencing contemporary music differ from their California peers who received other types of musical instruction. Finally, some information was gained as to the value of several innovative programs whose students composed the sample for this test-development enterprise and, conversely, the value of a test of aesthetic judgments in the assessment of musicality at the elementary school level.

In measuring achievement of Manhattanville and California students, these observations were made:

1. Musical children of the same age score similarly though drawn from divergent programs.

2. Sub-test scores--Melody, Harmony, Rhythm, Form, and Musical Style--differentiate between groups, reveal curricular emphases and the hierarchical nature of aesthetic sensitivity.

3. A growth gradient exists for test scores from grade 4 upward.

4. Manhattanville subjects were more perceptive than their California counterparts but were less likely to agree that the original composition excels its mutilation.

5. Mere perception scores (same or different) do not correlate as highly with teachers' ratings as do preference scores.

6. Individual differences are greater within samples than between them.

CONTENTS

	PAGE
INTRODUCTION	1
The Problem of Evaluation in the Manhattanville Music Curriculum Program	2
A Review of Objectives in Music Education Relevant to This Investigation	3
Purpose of the Study	6
Limitations	7
Hypotheses	11
Organization of the Report	11
RELATED RESEARCH LEADING TO THE DEVELOPMENT OF THE CALIFORNIA TEST OF AESTHETIC JUDGMENTS IN MUSIC	14
Evaluation of Young Audience Programs for Primary Grade Children	14
Madera PACE Project: Developing Musicality in the Elementary Schools	20
San Diego PACE Project: Impetus for the Creative Arts	27
RESULTS OF THE INVESTIGATION	31
Source of Data	31
Evaluating Selected Tests for Statewide Music Testing in California	32
Description of the Tests Studied	33

PAGE

Methodology 41

Statistical Treatment of the Data 42

Measurement of Aesthetic Sensitivity in the Manhattanville Music Curriculum Program 52

DISCUSSION 57

The Value of Aesthetic Judgments in the Assessment of Musical Achievement 57

Evaluation of the Manhattanville Music Curriculum Program 60

Summary 63

APPENDIX 68

Description of the California Test of Aesthetic Judgments in Music 69

LIST OF TABLES AND CHARTS

TABLE		PAGE
I	Charts Illustrating the Hierarchies of Learning	17
II	Results of the Testing Program Analysis of Covariance	18
III	Correlation Between Teachers' Ratings and Test Data	25
IV	Comparison of Experimental and Control Samples by Grades	26
V	Growth in Aesthetic Sensitivity, San Diego PACE Project, "Impetus for the Creative Arts"	30
VI	Tests and Subjects in California Evaluation	41
VII	Reliability of the Tests Investigated	43
VIII	Validity of the Tests	44
IX	Summary of Means, Standard Deviation, t Tests and Intercorrelations (Fall Testing)	45
X	Evidence of Growth in Musicality (Fall and Spring Testing)	46
XI	Growth in Aesthetic Sensitivity – Music Specialist Versus Classroom Teacher	47
XII	Results of the Testing, Manhattanville Music Curriculum Program	53
XIII	Comparison of Aesthetic Judgment Scores of Students Participating in California PACE Projects with Those of the Manhattanville Program (Post-Instructional Testing, California Test of Aesthetic Judgments in Music)	54

TABLE PAGE

XIV Percent Passing Items in Each Catetory. 55

XV Comparison of First and Second Year Test Scores for Primary Grade Students in the Madera "Orff" Project 59

CHARTS PAGE

Comparison of Attitude Scores of High and Low Achievers on California Test of Aesthetic Judgments 49

Comparison of Experimental and Control Schools on Responses to Questionnaire 51

CHAPTER I

INTRODUCTION

"To enable the right pupils to receive the right education from the right teachers at the right time," Remmers and Gage[1] judiciously observed, "may be considered the aim of every program of evaluation." This statement suggests that teachers must forever evaluate and re-evaluate the achievement of their students to discover whether the learning processes through which they propose to lead them will bear fruit for each student individually and for the society in which he lives.

Musicians have long appreciated the value of systematic evaluation in music education. The phenomenal growth of the music competition-festivals witnesses the consensus among musicians that assessment is indeed the second side of the coin of which teaching is the first. Evaluation, however, has almost been totally confined to the area of musical performance. Little or no thought has been given to measurement of conceptual development or aesthetic sensitivity of children who participate in music education throughout the elementary school solely through courses described as "General Music." Current attention to "cost-effect models" and to educational objectives defined in behavioral terms, emphasizes the need for evaluation in all music education classes. The crux of the problem is this: musicians have little knowledge of, or faith in, the battery of tests commercially available which purport to measure what is being taught. This research hopes to bring some light upon the matter.

The increased realization of the importance of describing objectives for music in behavioral terms, parallels the increased awareness of the value of evaluation. Interest in expressing goals in behavioral terms is a recognition that one can know the effectiveness of his educational endeavors only to the extent that he can

---

[1] Remmers, H. E. and Gage, N. L., Educational Measurement and Evaluation (New York: Harpers, 1943), p. 1.

observe changes in human behavior brought about by these endeavors. The vitality of objectives depends on evaluation. Objectives are standards by which to measure, but when careful, objective measurement does not take place, they are little more than superfluous gratuities. Attempts to improve the programs of music education, leading to the development of concepts, skills, and appreciations, can be based on nothing more than pedagogic intuition unless there is an effort made to determine accurately what children are learning in music classes. This is what systematic evaluation can do.

Need for evaluation is equally important for the classroom music teacher in the most orthodox or provincial of music programs and for those in leadership roles who would expand the frontiers of music curriculum through imaginative, innovative programs. One of the purposes of this study into the nature of aesthetic sensitivity and its measurement was to develop an instrument of evaluation for one such program, the Manhattanville Music Curriculum Program.

The Problem of Evaluation in the Manhattanville Music Curriculum Program

The Manhattanville Music Curriculum Program is an extensive and highly unified assault on the problems of music learning. Aligned psychologically and structurally with similar "new education" projects in other fields of learning, the MMCP is concentrating on the development of a music study and educational procedures which, while valid in terms of the nature of the art, allow the student to become intrinsically involved and discover the concepts of music through his own creative exploration.

Financed by the Arts and Humanities Division of the U. S. O. E., the entire program spans a four-year period of research, experimentation, pilot study, and demonstration activity. Included among the forty-seven active contributors and experimenters working on the project are music educators, composers, musicologists, and experts in the field of cognitive science. In addition, over one hundred eighty music educators are currently preparing to join the MMCP team as educational experimenters.

The program, under the direction of Ronald B. Thomas, is involved in experimental activity at all levels of education from the primary grades through college. In order to leave unfettered its daring and extensive format, the Manhattanville Program preferred not to include a firm statement of its expectancies and a scheme for

their evaluation. Two objectives were inherent in the structure of the program, however: 1) to develop aesthetic sensitivity to music, and 2) to prevent closure, i. e. , to insure open-mindedness towards new sounds in music as well as towards traditional music in our heritage.

"In curriculum development, " Thomas writes, "our first problem is to search for relevance - environmental, artistic, and personal, and construct a program which is geared to those relevant factors which we can even crudely establish. Of course, there is a great deal of supposition in this process, but there is little hard material that can guide us. Our second step is to assess what we have accomplished by striving to know what situations and information the student has acquired and stored for his own personal use. We must also discover how the student gained this information, and how he has structured it so that it has become usable to him. With this knowledge we can begin to understand the problems that stand between the student and accomplishment, and we can determine the experience and information which is vital to him. Only after these steps can we realistically identify objectives and construct a priority of objectives which have significance to both the student and the art. "[1]

The initial concern of this study, therefore, was to identify those musical behaviors which represent a consensus of the objectives of music education in the elementary schools of the Manhattanville Program and which might lend themselves to objective measurements. The major thrust of the research thereafter was to develop an instrument for evaluation of these behaviors.

A Review of Objectives in Music Education Relevant to This Investigation

The Yale Seminar on Music Education described the purpose of music education succinctly: "The purpose of music education throughout the grades is to develop musicality!"[2] Musicality was then

---

[1] From correspondence to the author from Ronald Thomas, December 13, 1968.

[2] Palisca, Claude V., Seminar on Music Education. Cooperative Research Project No. G-013, (Yale University, New Haven, Conn. , 1963), p. 43.

defined as the ability to express through time and pitch the mental image of a musical idea. Conversely, it is the ability to grasp in its entirety a musical idea which has been expressed. This latter definition of musicality is supportive of the view expressed in the Framework for Music in California Schools that music education ought indeed to be aesthetic education. The word "aesthetic" in its Greek origin, aesthetes, means "one who perceives."

In discussing the changing goals of music education at the Tanglewood Symposium,[1] Broudy expressed the view that music education should develop connoisseurship. "Connoisseurship or 'enlightened cherishing' requires aesthetic judgment, for it is in aesthetic judgment that we not only know what we like, but can also make some attempt at giving reasons for liking it." Broudy would characterize the behavior of an aesthetic person in demonstrable terms as one who can make aesthetic judgments in music.

Kaplan[2] also emphasized this viewpoint in the following statement:

> Music teaching today aims at developing a knowledgeable audience. While liking or disliking a piece of music is the prerogative of the student, his decision should be based on knowledge of the music. The aim of the teacher is to place the student in contact with the objects, the skills, the capacity for response, and the factual, measurable content of musical works. Music education thus should help students base their musical choices on defensible values. In any music curriculum concerned with values, teaching for aesthetic judgment is a fundamental goal.

Other educators have expressed a variety of rationales for including music in the curriculum of public education, and in many instances have justified music's place in the curriculum because of its contribution to non-musical goals. This faulty reasoning has affected music education at both the elementary and secondary levels, and has been equally short-sighted at each. McMurray and Broudy criticized this point of view as follows:

---

[1] Choate, Robert. Tanglewood Symposium, "Music in American Society," (MENC, Chicago, 1967), p. 22.

[2] Ibid., p. 22.

What is at least potentially bad about this way of thinking is that it conceives music as an instrument to the realization of non-musical values. It recognizes nothing distinctive and unique in music experience itself and claims for music only that it helps pupils develop other talents and learnings. [1]

. . . It is safer, therefore, to think of music in general education as the cultivation of capacities for realizing value. [2]

Music education is justified because, when the more refined portions of our musical culture are communicated, the person to whom they are communicated will find in music what he would not have been able to find otherwise, thereby expanding his environment and increasing his power to find a good life through deliberate guidance of his behavior and its outcomes. In a world where patterns of sound are omnipresent, he will have increased power to control what happens to him musically, and to make the aesthetic quality of his experience less a matter of mere accident. This is an important kind of contribution to anyone's general education, and it respects the values of musical experience as something other than minor instrumentalities to nonmusical goals. [3]

In defining the aims of music education for the Manhattanville Music Curriculum Program, Thomas[4] writes as follows:

---

[1]McMurray, Foster, "Pragmatism in Music Education," Chapter II, Basic Concepts in Music Education, The Fifty-seventh Yearbook of the National Society for the Study of Education, Part I (Chicago: University of Chicago Press, 1958), p. 38.

[2]Choate, op. cit., p. 76.

[3]McMurray, op. cit., p. 42.

[4]Thomas, Ronald, MMCP Synthesis 1969 (Manhattanville College, Purchase, New York), p. 8.

The ultimate achievement of musicality implies knowledge, attitudes, and skills, but it goes beyond these objectives. It includes the awakening of an aesthetic sense, the ability to comprehend beauty and to find meaning on a plane beyond analysis of mechanics, techniques, or even concepts. The capacity for such feelings involves more than the intellect and the senses. It involves the emotions and the spirit of the individual. Through a program which emphasizes individual analysis, judgment, and creative thinking in all experiences, the development of aesthetic sensitivity must be considered to be an individual experience. Aesthetic insight is a condition which can exist only in terms of one's personal affinity to the nature of the art. Consequently, it can only be an intimate response. The conditions for this response can be fostered by educational strategies which demand analytical, judicial, and creative thinking. While the objective, aesthetic sensitivity cannot be directly communicated, the conditions required to meet this educational aim are provided in the other objectives and processes in the curriculum.

Purpose of the Study

It was the purpose of this research to study growth in aesthetic sensitivity in music and to investigate the value of aesthetic judgments in the assessment of musicality of elementary school children. To do so it was necessary to examine and reconstruct instruments of evaluation which purport to measure aesthetic sensitivity through the ability to make judgments of the appropriate usage of the materials of music: melody, harmony, rhythm, and timbre as they affect form and musical style. These judgments are held to be evidential of the ability to "grasp in their entirety musical ideas heard," as the Yale Seminar on Music Education defined musicality so aptly.

A concomitant purpose of this enterprise was to compare the aesthetic responses of children in several exemplary music programs, including the Manhattanville Music Curriculum Project, to determine what aesthetic responses typically may be expected, to determine in which ways those students rated as being successful by their respective teachers differ in musical sensitivity from those

classmates assessed as being least musical, and finally, the effect of various types of instruction upon the rate of development of aesthetic sensitivity to music. It was not the thesis of this investigation that one should expect the same "correct" response for each judgmental task by all children at all ages. Such a position suggests that tastes are static, unresponsive to training. It is preferred, rather, to define musicality merely as the ability to adjust to the demands of the medium. Musicality, by definition, is merely a word used in this study to describe human behavior from a particular point of view--namely, that of a musician's. In this research, musicality describes specifically the behavior which marks the musically successful elementary school students as identified by their music teachers.

The California Framework for Music Education describes several behaviors comprising the aesthetic experience which tend to clarify the criterion variable to be used in this investigation. The two most important are perceiving and reacting. Evaluation is a sub-behavior to perceiving and reacting in the aesthetic experience. It is characterized by the ability to make aesthetic judgments in music. Rightfully, it should be considered as a means to the end of being able to enjoy music with more sophistication, more perspective, more keenness of mind and of feeling. It represents the individual's realization of value. As such, it represents a summation of the purposes of music education as aesthetic education. It is appropriate, then, for this research to direct its focus upon this manifestation of the aesthetic experience.

Limitations

An important limitation imposed upon those studying aesthetic judgment lies in the definition of, or assumption of, absolute standards of goodness. Hevner[1] assumes that the creation of an artist, whose work is generally acclaimed for its merit, is beautiful, and that it is more beautiful than the same creation altered by a deliberate attempt to mutilate its various beautiful qualities. She believes beautiful music is that music which is played and enjoyed by the best professional musicians and that, obviously, the selection of "good" music on the basis of the majority vote would be

---

[1]Hevner, K., "Tests for Appreciation of Music," University of Oregon Publications, IV (1934) No. 6, p. 111.

a serious error. She argues that one might as well decide that a four-room, frame house is the ideal place to live because the majority of American families live in houses of that type. The judgment of the man on the street is not the measure of musical values. Music is kept alive, rather, by the passionate few.

Shoen[2] defines a musically "good" melody as one that creates a complete impression just as does a properly constructed sentence. There are rules, of course, but a musically-minded person has a feeling for total sequences that is active before knowledge of laws or reason is obtained, just as a child grows up with a feeling for the fitness or appropriateness of the words of his native language long before he is taught grammar.

Although recognizing a need for the empirical method of validating the items in his test of musical ability, Wing[2] presupposed certain immutable standards in choosing his musical material, as evidenced by this statement: "If a piece of music (a) were taken from the best works of expert composers, (b) were thought good by the consensus of musicians, (c) had survived the test of time, (d) were from a standard edition, it was assumed that it would be representative of <u>good</u> art."

Farnsworth[3] points out that history discloses change in musical standards. The taste of Western culture has changed appreciably from time to time. Leading composers have been non-conformists. Although they study the musical traditions of their period, they almost invariably adapt the rules to fit their own needs. Thus the rules as expounded in the manuals do not remain static, but, rather, are frequently recodified. As an illustration of our changing preferences, one might take the situation of the definite keynote. Although there existed from early days the "final" of the Ambrosian and Gregorian modes, the true key tone, or tonic, apparently did not enter European music until the thirteenth century.

---

[1]Schoen, Max. "Tests of Musical Feeling and Understanding," <u>Journal of Comparative Psychology</u>, V (1925), p. 270.

[2]Wing, Herbert, "Tests of Musical Ability and Appreciation," <u>British Journal of Psychology</u>, Monograph Supplement XXVII (London, Cambridge University Press, 1948), p. 2.

[3]Farnsworth, Paul. "Musical Taste, Its Measurement and Cultural Nature," <u>Educational Psychology</u> (Stanford University Publication, Vol. II, No. 1), p. 23.

It later achieved such a strangle hold on musical practice that in 1893 one authority expressed the view that without a clearly defined tonality music is impossible. Yet during this same period of the tonic's extreme popularity, people of other cultures were enjoying keyless music, and now polytonal and atonal compositions are commonplace in our own culture. For another example of taste change, attention can be called to parallel movement, which was so much relished at one time, and then was sternly banned. Now it has reappeared and at times has achieved a desirable status.

The hypothesis that contemporary taste in music is in a large measure culturally derived can be demonstrated through the data of anthropology. It has been shown that the Western World love for simple rhythms, careful tuning, sized tonal steps, harmonies, and the tonic effect is not shared the world over. The African's sensitivity for complicated rhythmic patterns was so far beyond the taste and perceptual abilities of many of the early missionaries that they commonly reported the African to be arhythmical. The Chinese often appear oblivious to mistunings; they love music which has no harmony in the Occidental sense of the word. Yet Orientals can learn to love Western music, and, indeed, with continued residence in America, come to appreciate our musical principles and gradually to develop facility in the perception of small auditory differences. When constantly subjected to poor tuning, the American loses his need for, and sensitivity to, pitch exactitude.

In developing the tests of aesthetic judgment, the underlying rationale of the investigator was that aesthetic tastes are culturally derived. It was recognized that tastes change and that a musical experience which is thought to be highly appropriate in terms of harmony, phrasing, or melodic line at one age may well be entirely scorned at another. Moreover, it was evident that in varied socioeconomic classes, one should not declare, a priori, the level of sophistication of musical taste each specific grade level ought to have reached. The tests, therefore, as developed, depended upon empirical validation. That is to say, items were sought which require aesthetic judgments in music to discover how the most musical children might respond. Only when such items differentiated between students rated as most musical by their teachers and those students rated as least musical, however, were the items thought suitable for further investigation into the nature of aesthetic sensitivity and the musical experiences which promote its development. Initially, the researchers were prepared to accept, as correct, any responses made by persons identified as being most musical. Interestingly, except for the most contemporary musical

examples, the most musical children did choose the original versions of all of the musical examples, that is, they preferred the music as composed by the "recognized" composers compared to a mutilated version. Only those items which significantly differentiated between high and low achievers, both as to the total test score and as rated by their teachers, were retained in the final test.

To illustrate this view further, in this project it was not assumed that fourth graders ought to be able to distinguish, aesthetically, between a harmonization of America by Carey and one by Stravinsky. Rather, the study proposed to find out whether fourth grade school children could, and if so, upon what basis the distinction was made.

Neither was it assumed that a piece of music is "good" because Beethoven wrote it or because it had stood the test of time. Rather, the study sought to determine if elementary school children are likely to prefer the published edition of Beethoven's Spring Violin Sonata over an earlier sketch found in his notebook. Beethoven undoubtedly preferred the published version. However, the intent of this research was to determine if that preference indeed marked the most musical children as they are exposed to this music in the schools. If aesthetic choices do distinguish the high achiever from the low achiever, either grade-wise or by musical achievement as judged by their teachers, such items would appear suitable for further investigation and possible use in evaluating various modes of instruction.

In summary, it is believed that the science of aesthetics is not an exact science but rather a normative discipline dealing with values. Beauty is indeed in the eye of the beholder! Items for the tests were chosen, therefore, on the basis that they differentiated between persons known to be musical and those observed to be less musical. They were not chosen particularly because of values inherent in the art works, though as observed, this may have been the case for all of the items except those of the contemporary composers.

To formalists such a definition of "standards of goodness" of music may appear to be untenable. Therefore, at the outset this view is described as a limitation of the study.

Hypotheses

It is postulated that aesthetic judgments may serve as the organizing factor of the elements of auditory imagery; and, therefore, students who score highest on standardized tests of aural perception and musical concepts will also score highest on a test of aesthetic judgments in music. Furthermore, scores on an aesthetic judgments in music test will correlate significantly higher with success in music, as determined by teachers' ratings, than will scores on mere perception tests.

A second hypothesis to be tested is that musical aesthetic judgments are culturally derived, i. e., subject to the educational process, and that, therefore, there will be a growth gradient of mean scores on a test of aesthetic judgment through the elementary grades which, in turn, will be significantly affected by the quality of the music programs in the schools.

A third hypothesis, in opposition to the second, will also be tested. The third hypothesis is that ability to make aesthetic judgments may reflect a quasi-musical "intelligence"; therefore, differences on test scores between the high and low achievers will be significantly greater within any single grade than will the mean scores between consecutive elementary school grades. Musical intelligence is here defined as the ability to adjust to the demands of the medium, the essence, of course, being the rate of adjustment.

If the first two hypotheses prove to be true, then it would appear that a test of aesthetic judgments in music might well serve as a much-needed instrument for measuring achievement in school music. Such a test would indeed prove useful in evaluating the effectiveness of various curricula purporting to develop musicality. If the third hypothesis is proven to be true, the test may have virtue in the early identification of the precocious but would not appear to reflect the effect of formal music education.

Organization of the Report

I. Related Reseach

It is customary for investigators undertaking curriculum research to review the research of others in the field. This is done for the dual purpose of assuring the reader that the research is not a reharrowing of an overly-cultivated field and, secondly, that the

researcher is in the position of acquaintance which will permit his use of previous findings in extending knowledge.

Inasmuch as the field of music evaluation has been so well summarized in recent publications (notably that of Lehman[1] and Whybrew[2]), and major research in music appreciation has been carefully drawn together by Hevner and by Wing, it would be redundant to engage in such analyses here. Consequently, the following chapter of this report will be devoted, instead, to research undertaken as related pilot studies necessary in developing the test required by this investigation into the nature of aesthetic sensitivity and its cultivation.. The succeeding chapter will then report the evaluation of the Manhattanville Music Curriculum Project which was a major component of this research proposal.

The population for the pilot studies which contributed to the development of the final test and its validation was comprised of the following sub-samples:

1. 400 second and third grade students from Oakland and Berkeley public schools participating in the San Francisco Bay Area Young Audience Program Evaluation;

2. 1800 children from 27 schools in the Madera, California, PACE project which was concurrently evaluating the Orff Techniques for Developing Musicality at the elementary school level;

3. 1600 children in the San Diego, California, PACE project, Impetus for the Creative Arts, which proposed to develop musicality through "Saturday Conservatory Training" for the gifted student and by enriching the classroom musical experience through the utilization of demonstration performances by professionals in the community.

Each investigation serves as an essential pilot study, yielding basic information concerning the nature of aesthetic sensitivity and

---

[1] Lehman, Paul R. Tests and Measurements in Music (Englewood Cliffs, N. J.: Prentice Hall, Inc., 1968), 99 p.

[2] Whybrew, William E. Measurement and Evaluation in Music (Dubuque, Iowa: Sm. C. Brown, Co., 1962), 184 p.

its cultivation in the public schools. For example, in the Young Audience Program Evaluation Project, which is reviewed, considerable evidence was obtained concerning the hierarchical nature of the conceptual development leading to aesthetic judgment. In the Madera PACE Project, noted for its exploratory Orff Techniques, an eclectic test drawn from existing music appreciation tests substantiated the logic of estimating musical growth through tests of aesthetic judgments. The San Diego Project, Impetus for the Creative Arts, permitted the construction of a new test drawn from relevant elementary school musical experiences as well as from the improvisational work of the Manhattanville Music Curriculum Program.

Although these research projects were independently conceived and carried out, they are believed to be most relevant and serve as pilot studies leading to the evaluation of the Manhattanville Music Curriculum Project.

II. Evaluation of Growth in Aesthetic Sensitivity of Students in the Manhattanville Music Curriculum Program

In the fall and spring of the school year 1968-69, the California Test of Aesthetic Judgments in Music was administered to 800 students in the fourth, fifth, sixth, seventh, and eighth grades selected from the schools participating in the Manhattanville Music Curriculum Program. Schools in Portland, Oregon; Denver, Colorado; San Angelo, Texas; Yonkers, New York; West Hartford, Connecticut; Cresskill, New Jersey; and Larchmont, New York comprised the sample. Where possible, control classes from these schools were also tested. Pre and post testing data were submitted to an analysis of variance.

III. Field Testing of the California Test of Aesthetic Judgments in Music as a Part of the California Department of Education Study into Evaluation of Music Education in the Elementary Schools of California.

The California Test of Aesthetic Judgments in Music was evaluated for possible state-wide usage by the Northern Section of the California Music Educators Association under a grant from the California State Department of Education. Data obtained from that investigation permit certain conclusions to be drawn as to the efficacy of the test in reference to other music achievement tests.

CHAPTER II

RELATED RESEARCH LEADING TO THE DEVELOPMENT OF THE CALIFORNIA TEST OF AESTHETIC JUDGMENTS IN MUSIC

Evaluation of Young Audience Music Programs for Primary Grade Children

The highly successful Young Audience Music Programs, which for more than a decade have favorably influenced music education in the elementary schools, have, for the most part, geared their presentation to the upper grades of the elementary school. This practice is prevalent for several reasons: (1) the narrative portions of the programs are suspected of being too sophisticated for younger minds; (2) the attention span of young children is notably shorter than their older schoolmates; (3) instrumental instruction, for which the young audience programs have served as a powerful motivator, are typically begun in the intermediate grades; (4) and, finally, there is simply a paucity of evidence to show that very young children can profit from the exposure to serious music in more than a superficial way. This research attempted to gather evidence as to the feasibility of introducing the Young Audience Programs to children in the 1st, 2nd, and 3rd grades. It was the result of an evaluation made of six programs as presented to primary grade children in schools of San Francisco, Oakland, Richmond, and Berkeley, California.

Video tapes of portions of six Young Audience Programs were made in the winter of 1968. From these excerpts a series of twenty-six scenes were selected and presented in audio-visual form by way of closed-circuit television to students comprising experimental and control samples. Questions were asked concerning the happenings on the tape.

As hypothesized, the cognitive tasks required of the testee were indeed found to be hierarchical in their format. In the concepts of pitch, for example, a woodwind quintet demonstration of the effects on pitch of shortening a soda straw on pitch - revealed that all students had developed the concept of highness and lowness in pitch.

The effect on pitch of emptying a Coke bottle as one blows across the top further demonstrated the relationship of the length of the resonating chamber and resulting pitch. From these demonstrations children were led to anticipate the sound of the longest of the woodwinds in the quintet, the bassoon. This, in turn, led to the definition of the word "bass" as meaning "low." This concept served as the basis for several additional tasks. For example, the task of remembering and generalizing the pitch concepts to the bass voice and bass recorder led finally to the making of an aesthetic judgment of the appropriateness of a bass or soprano voice for singing the "Bell Song" from Lakme. This final task represented the highest order of learning or generalization of experience.

Examples of tasks required in the test.

Pitch--recognize that:

1. The pitch changes as the soda straw is shortened.
2. A coke bottle sounds lower as it is emptied.
3. Pitch is related to the length of the pipe. (The bassoon is the bass of the woodwind family.)
4. A bass krummhorn sounds like the low register of the bassoon - a bass recorder is a bass instrument.
5. The viola should not sound higher than a violin.
6. The "Bell Song" from Lakme is more appropriate for the soprano voice than the bass voice.

Rhythm--recognize that:

1. Each verse of the ballad is accompanied by a different rhythmical pattern by the small drum.
2. The underlying beat remains the same.
3. The tempo remains constant for the first several verses, but is slower for the final verse.
4. Excitement can be generated by varying the rhythmical accompaniment even if the tempo is constant.
5. Changing the tempo does alter the mood effect. Therefore, the song ends "sadly." (A judgment of appropriateness of tempo.)

Timbre--recognize that:

1. A flute sounds like a flute, an oboe differs from a bass clarinet, and a viola does not sound like a French horn.

2. A krummhorn sounds more like a bassoon than it does like a flute or French horn.
3. A krummhorn is a double reed instrument. The oboe and bassoon are also double reed instruments.
4. Rebec and cello belong to the same family (bowed, stringed instruments).
5. Pitch and timbre are not equivalent. A soprano recorder does have a kinship to the alto, tenor, and bass recorder. A tuba does not belong to this family.
6. A recorder is more appropriate to depict birds singing in spring than a rebec.
7. Jazz Pizzicato played pizzicato is more appropriate in style than is a performance using bows.

Form--recognize that:

1. Copy-cat music (canon) is imitative music.
2. A rondo is like a song with many different stanzas with the same chorus coming in between each stanza.
3. A rondo, with its recurring theme, is more interesting to listen to than a piece comprised solely of its episodes.

TABLE I

Charts Illustrating the Hierarchies of Learning

I PITCH

Percent of second graders passing each successive test

100% 80% 60% 30% 25%

25% made the aesthetic judgment concerning the appropriateness of bass or soprano voice for "Bell Song" from Lakme.

37% generalized this concept to know that a viola, larger than a violin, should not sound higher than the violin.

60% learned the meaning of bass as being low and were able to categorize a bass recorder sound.

78% learned the effect of lengthening the pipe on pitch; the bassoon should sound lower than other woodwind instruments; recognized higher and lower sounds.

100% recognized that the pitch changed as the soda straw was shortened.

II TIMBRE

100% 50% 40%

40% preferred the use of a recorder for a "Spring Song" over the rebec.

44% generalized the fact that a rebec is the grandfather of modern bowed instruments.

47% knew that the French horn does not sound like a viola.

50% classified the bassoon as sounding like the krummhorn, and therefore was a double-reed instrument.

96% agreed that a flute does not sound like a double-reed instrument.

100% recognized that the tones produced by different instruments are different from each other.

TABLE II

Results of the Testing Program Analysis of Covariance

	Pre-instruction Mean	Adjusted Post-Instr. Mean	Degree of Freedom	f Val.	Level of Sign.
Total 2nd Grade Experimental	10. 11	11. 79	226	18. 11	. 01
Control	6. 18	9. 64			
Total 3rd Grade Experimental	12. 04	15. 37	148	4. 71	. 05
Control	12. 82	14. 75			

Interpretation of the Results

From Table II it may be seen that when the control and experimental samples are statistically equated, that is, consideration is given to the regression line to yield an adjusted final mean, the growth in musical sensitivity is significantly greater for the experimental groups (the groups receiving the Young Audience Programs).

The results of the testing program clearly indicate that the Young Audience Programs have brought about a significant development in the concepts of pitch, rhythm, and timbre for the primary grade children in this evaluation program. Whether the increase in scores from pre-exposure testing to the final testing may be due to conceptual development inherent in the programs or whether the results were highly influenced by the obvious attentive attitude and interest of the experimental sample cannot clearly be ascertained. In aesthetic education, as elsewhere, cognitive development does not proceed independent of the affective. Many music teachers would argue that how second grade children feel towards music is probably as important as what they know about it! Fortunately, the results of the testing program indicate a favorable combination of the cognitive and affective domains for growth in musical perception.

The fact that one of the control samples also made significant gains between first and second testing suggests that this test itself induced learning. This phenomenon was not unexpected,

inasmuch as the test was constructed of scenes drawn directly from the Young Audience Programs. On the other hand, the failure of the third grade control sample to make gains implies that even a video tape is a poor substitute for live performance. It would appear that student rapport with "live music" is desirable for the greatest accomplishment to be attained. While this pilot research indicates that the task of making aesthetic judgments is too much to expect of the second grade student, it does reveal the hierarchical nature of aesthetic sensitivity as directed towards the higher problem-solving task in music, that of making aesthetic judgments.

The question pertinent to aesthetic research is this: After a student has learned to distinguish between a flute and a krummhorn is he better able to "appreciate" music than his friends who cannot? Since we declare that enlightened cherishing or "connoisseurship" is the goal of music education as aesthetic education, a case should be made for knowledge. Isn't an educated taste better than a raw one? Before a student can be expected to make judgments of appropriateness of the use of the elements of music must not he be perceptive of these elements?

For the second grade student, an increased awareness of sound qualities, pitch relationships, and rhythms in music is prerequisite to judgment concerning their appropriate usage. The hierarchical nature of the task involved in making aesthetic judgments suggests that music education as aesthetic education must provide additional experiences to bring about growth in the affective or judgmental domain. This measurement tended to show the effectiveness of Young Audience Programs in bringing about increased sensitivity to the elements of music. Furthermore, the evaluation went beyond mere rote memory and required that the second grade students generalize their experience. They were led by the programs to perceive and see relationships, i.e., to conceptualize, but the students could not make appropriate aesthetic judgments! Perhaps in this period of growth, which Piaget calls the concrete operational, responding abstractly to judgmental tasks is beyong reasonable expectation. The following research account of the Madera PACE Project tests this assumption with a second type of aesthetic judgments test. This eclectic test of aesthetic judgments in music was one made of the most efficient items of the available commercial music appreciation tests.

Madera PACE Project: Developing Musicality in the Elementary Schools

It is not uncommon in an innovative program such as the Madera County PACE Project, Developing Musicality in the Elementary School, to direct emphasis solely towards attitudinal changes. Evaluation, then, tends towards the chronology of changes in children's attitudes toward music, the increased teacher participation, the number of demonstrations performed for visiting dignitaries, and so on.

The Madera project is particularly relevant to our research in that, by definition, it demanded a harsher appraisal of its outcomes. It required a measurement of changes in student behavior observed from the Yale Seminar's point of view, i.e., growth in musicality.

Although musicality is defined as "the ability to express a musical idea in its completeness, or, conversely, the ability to comprehend a musical idea, in its entirety, which has been expressed," the difficulty of measuring, fairly, the ability to express musical ideas, through instrumental and vocal performance, through composition and dance, caused a necessary delimitation of the proposed measurement to include merely the converse definition of musicality--"the ability to comprehend musical ideas." The necessity of the delimitation is obvious. The inclusion of the term "the ability to express musical ideas" demands equation of the varied means of expression as to make the testing of the hypothesis almost totally dependent upon the means of expressing musical ideas. On the other hand, the ability to understand music is an accepted outcome of the various avenues of music education and more readily lends itself to measurement. The stringent measure of this latter capacity, it was thought, should be a measure of aesthetic sensitivity in music as reflected by the ability to make aesthetic judgments. Using this definition the evaluator set about to construct a test eclectically from available music appreciation tests which purport to measure achievement in musicality--defined as the ability to comprehend musical ideas. The test was a composite of aesthetic judgments tests in music which, by measuring the subject's ability to make judgments of the appropriateness of the usage of the elements of music, melody, harmony, rhythm, form, and timbre, evaluates his ability to comprehend, in their entirety, musical ideas which have been expressed.

Primary sources for materials for such a test were the music "talent" tests which, too, have been validated on teachers' ratings--particularly those batteries which used aesthetic sensitivity

measures to predict success in music. The Wing Test of Musical Intelligence, the Gordon Musical Profile, the Kwalwasser-Dykema Test of Melodic Taste, and the Kyme Test of Aesthetic Judgments in Music were immediately available as preliminary measures. The Oregon Test of Music Appreciation by Hevner and Landsbury, though never intended to be a test for elementary school children, served as a model for the aesthetic judgments test. It follows the Hevner design in that the items require the testee to make a judgment of appropriateness of the harmony, rhythm, or melody of the original form of a composition when compared to an alternate version which had been mutilated in terms of this essence.

Description of the Tests Used as a Preliminary Battery

The Hevner-Landsbury Tests of Musical Appreciation attempts to measure three aspects of music appreciation: music discrimination, musical concepts, and attitude toward music.

The discrimination test was designed to measure judgment of beauty and appropriateness in music. It consists of forty-eight items taken from various compositions and is presented by means of piano performances on recordings. Each item is presented one time in its original version and once in a mutilated version. Mutilations are achieved in a variety of ways, including the extension, reduction, and alteration of rhythm, melody, harmony, and form. The subject is required first to judge which of the two is the superior version and, secondly, to judge whether the mutilation involved the harmony, melody, or rhythm. Hevner used a weighted scoring method inasmuch as no credit is given for the correct identification of the type of mutilation unless the subject has also selected the original version as the better of the two. Correlation with training in music and test scores was .64; correlation with the Seashore battery of tests was .54.

The musical concept tests is designed to measure comprehension and understanding of musical composition as a whole. A series of questions and statements, both affective and technical in nature, is given concerning nine classical compositions. True-false answers are scored, based on agreement with expert opinion.

The test of attitude towards music follows the Thurston method of scale construction. Five statements about music are arranged on an eleven-point scale and the subject's score is determined on the basis of the scale values which he has checked.

The Kwalwasser-Dykema Tests. This test consists of a battery of ten tests given by means of recordings. The tests differ from the Seashore battery in that they use actual musical material. For this reason, the Kwalwasser-Dykema Tests are more acceptable to musicians. It was chosen for this study because part of the test is concerned with aesthetic sensitivity in music.

The tonal memory test involves 25 pairs of tonal patterns which become increasingly complex. The subject must decide whether the two patterns of each pair are the same or different.

The quality discrimination test involves 30 items in which a melodic fragment is played twice, either by the same instrument or by a different one. The subject must decide whether the tonal quality is the same in each instance.

The intensity discrimination test consists of 30 items; 15 pairs of tone, and 15 pairs of chords. Intensity of sound is the only variable. The subject must judge whether the second is softer or louder than the first.

The tonal movement test presents 30 melodic patterns of four tones each. The patterns are incomplete and require a fifth tone for satisfactory aesthetic completion. The subject must decide whether the fifth tone should be below the fourth tone of the pattern or above it.

In the time discrimination test, three tones are heard, with the first and third of equal duration but with the second tone variable. The subject must decide whether the three tones are of equal or different duration. There are 25 items in this part of the test.

In the rhythmic discrimination test, 25 pairs of rhythmic patterns are offered in which differences of intensity and duration, or both, exist. The subject must decide whether the two patterns are the same or different.

In the pitch discrimination test, fourth tones are heard, each being sustained for three seconds. On some of the tones a fluctuation in pitch takes place. The subject must identify the tones in which the change has taken place.

In the melodic taste test, 10 items are given twice (totaling 20 trials in all). In the first phrase of each pair, the melodies are the same, but two second phrases are offered. The subject must decide which of the two second phrases is appropriate on the basis of congruity with the first phrase.

In the pitch imagery test, 25 tonal patterns are given in musical notation. The subject must decide whether the notation is the same as the pattern heard on the recording, or different.

The final test, the rhythmic imagery test, duplicates the technique of the pitch imagery test. Twenty-five rhythmic patterns are presented in notation and the subject must compare them with the pattern sounded on the recording.

The Kyme Test of Aesthetic Judgments in Music. Since the Kyme test will constitute a portion of the test used for this investigation, a review of its development may be appropriate.

This test consists of 53 paired performances in which the subject is asked to make judgments about the relative merit of each recorded performance heard. The present test is the piano version of a similar test of aesthetic judgments in music which consisted of the evaluation of paired performances, some taken from commercial recordings and others recorded at the Northern California Music Festivals. The Kyme test was empirically validated by retaining only those items in the test battery which differentiated between persons adjudged to be musical or less musical. As a validating sample, the upper and lower quartiles of 1,425 high school band and orchestra members in the San Francisco Bay Area were employed. The musicality of these students had been rated by their teachers who had observed the subjects in many musical situations for periods of one year or more. Those items which 75 percent of the most musical students agreed upon but which not more than 50 percent of the less musical persons were able similarly to assess were retained for the final test. The reliability of this test of 53 items, determined by the Spearman-Brown formula, was .80.

The Kyme Test of Aesthetic Judgments was administered to 412 ninth grade music students. When the test was scored as a test of aesthetic judgment, which required the organization of the elements of auditory imagery--pitch, intensity, timbre, and duration--into meaningful wholes, the mean correlation between test scores and teachers' ratings was computed to be .74. When the test was treated as a simple discrimination test that required the mere detection of differences between two performances which comprised each item, the correlation between test scores and the same teacher's ratings was found to be but .09. These results show that aesthetic judgments may be of some value in the assessment of musical capacity. It also suggests the limitations of tests which merely ask for the subject to make "same" or "different" responses, without requiring the organization of the elements of auditory imagery into meaningful Gestalten.

The Wing Tests of Musical Ability and Appreciation. In constructing this test, Wing preferred not to assume that certain given attributes make up musical ability but, rather, preferred to derive items from many tests, particularly those suggested by musicians, and then to delete those which failed to differentiate between persons known to be musical or unmusical. Through

statistical analysis, Wing selected the most promising course to follow. The final battery consists of six tests: (1) chord analysis, in which the subject is required to detect the number of notes played in a single chord; (2) pitch change, wherein the subject is to detect a change of notes in a short piece; (3) musical memory, in which the subject is to detect and identify by number the note changed in the second playing of short tunes; (4) rhythm accent, wherein the subject is to judge the better rhythmic accent in two performances of the same piece; (5) harmony, in which the subject is to judge the more appropriate of two harmonizations, using the same melody; (6) phrasing, wherein the subject is to judge the more appropriate phrasing in two performances of the same music. The complete test is recorded and may be used for subjects eight years old through adulthood.

The results of Wing's analysis disclosed two significant factors: (a) judgment of the appropriate musical arrangement, and (b) perceptive change in melody, chord, or number of notes. The first factor is responsible for more of the total variance than all of the other factors combined.

The Gordon Musical Aptitude Profile is classified into three main divisions: Tonal Imagery, Rhythm Imagery, and Musical Sensitivity. Each of these divisions is further subdivided into sections.

In describing the third test, Musical Sensitivity, which contains measures of musical preference, Gordon states that this division contributes a more comprehensive appraisal of basic musical aptitude, since musical creativity and expression are at least as important to success in music education as the ability to perceive tonal and rhythmic relationships among notes in a musical phrase.

A pilot study was made with 204 ninth grade students on the premise that teachers' ratings, which will be used as the external criterion of validity, would be more reliable if these ratings were based upon a long acquaintanceship with the students in several musical situations. As might be expected, there was a rather low correlation between teachers' ratings and test scores because the ninth grade students in music represent a stratified sample that is limited to those who elect to take music as a subject. Naturally, this lack of normal distribution of test scores had a negative effect on correlations. Table III shows the correlation between teachers' ratings and the various test scores.

TABLE III

Correlation between Teachers' Ratings and Test Data [1]

Data Sources	.r
Kwalwasser-Dykema	.006
Wing Test IV--Harmony	.097
Wing Test V--Phrasing	.034
Kyme Contemporary Music (piano items)	.203
Kyme Phrasing (instrumental ensembles)	.115
Gordon Melody	.128
Gordon Rhythm	.051
Gordon Musical Sensitivity	.194
Total Gordon Music Profile	.164
Hevner Melodic Items	.222
Hevner Rhythm Items	.180
Hevner Harmony Items	.024
Hevner Form Items	.141
Total Hevner-Landsbury Test of Music Appreciation	.233
Age	.390
Sex	.170
Years of Private Instruction	.327
Student Teacher Ranking of Subjects	.640

The most effective was a combination of the Hevner-Kyme Test measuring the appreciation of melody, harmony, rhythm, and form. Trial tests composed of 65 Hevner Music Appreciation Test items and Kyme Aesthetic Judgments in Music Test items were administered to children in the Madera County schools in the spring quarter 1967 in anticipation of the following year in which the innovational program was to be instituted. From this preliminary testing an eclectic test was derived which served as a measure of musical growth for the PACE experiment in Madera County. The composite test was administered to the Madera County students in September, 1967, and May, 1968.

---

[1] These data are reported in the Kyme Study "Developing Musicality at the Junior High School Level and the Contribution of Musical Composition to this Development," H.E.W. No. 2, 1968.

Table IV shows that the total experimental sample, analyzed grade by grade, is significantly different from the total control sample in three of the five grades.

TABLE IV

Comparison of Control and Experimental Samples by Grades

65 Item Test (First Year)

Total	N	Pre-Instr. Mean	Post-Instr. Mean	Adjusted Mean	Standard Error Adjusted Mean	F Value	Significance Level
Grade II							
Control	125	23.55	22.18	22.15	.53	48.40	.01
Exper.		23.47	27.18	27.83	.61		
Grade III							
Control	182	22.60	19.62	19.29	.26	13.09	.01
Exper.		20.90	20.62	21.04	.40		
Grade IV							
Control	240	21.45	23.85	24.12	.27	1.47	Not Signif.
Exper.		22.98	25.54	24.77	.46		
Grade V							
Control	186	22.31	26.18	26.51	.20	.02	Not Signif.
Exper.		24.87	28.09	26.43	.46		
Grade VI							
Control	34	23.60	25.91	26.01	.48	56.51	.01
Exper.		24.00	30.63	30.41	.33		

There was considerable basis for the expectation that a test made up of the most efficient items of music appreciation tests, which in turn have been validated by teachers' ratings of

success in music, would indeed reflect some change in behavior due to the varying degrees of efficiency of instruction. It is remarkable that the consistency of this differentiation should reach down as low as the second grade, for the test was primarily constructed for and validated upon the intermediate grades.

It is obvious any test measures more and less than it purports to measure. The correlation coefficient of .497 of test scores and teachers' estimates of success would suggest that the test is at least partially measuring what it purports to measure; on the other hand, it is measuring more than musicality. For example, it is measuring the willingness of the student to take the test, his ability to concentrate for a rather long period of time, and finally, it probably reflects the attitude of the student towards the subject of music in general, which, of course, serves as a motivator possibly accounting for high scores on the test as well as a regression on the part of the youngest children in the control sample due to taking a long test twice.

In the second year of the Madera Project it seemed advisable to shorten the test and to restructure some of the items to make them more suitable for very young children. It was thought advantageous, also, to include items of a fifth dimension, the assessment of suitable tone quality as a supplement to the four dimensions, melody, harmony, rhythm, and form, currently being assessed.

The test given in the second year - the California Test of Aesthetic Judgments, which is described in the appendix, incorporated these suggestions. The results are found in Table XII in comparison with the Manhattanville data. It must be observed that the test is apparently too difficult for second and third grade students, or the skill of using an IBM sheet for machine scoring needs to be rehearsed in advance of the testing. The mean scores of these grades are not significantly different from scores obtained through chance alone.

San Diego Project: Impetus for the Creative Arts

It is a rare occasion in developmental research that the investigator has a chance to re-examine his efforts. This research was considerably enhanced by an invitation by the San Diego County PACE Project, Impetus for the Creative Arts, to replicate the

evaluation strategy of the Madera project with some profit from that experience.

The PACE project originating in Chula Vista, San Diego County, was evaluated much in the same way as the Orff Project in Madera County. As in the Madera project, it was the purpose of the evaluation to compare growth in musicality, defined specifically as the ability to comprehend musical ideas. It utilized children in three schools in the San Diego District who received normal musical experiences in the elementary schools and three equated schools in the district who received musical experience provided by the PACE project, Impetus for the Creative Arts. A further evaluation was made of children who participated in the Saturday Conservatory program. It was possible in this evaluation to re-examine and refine the Madera composite Test of Aesthetic Judgments and to develop additional items more suitable for the elementary school. The criticism that a test composed totally of piano items was boring to young children was taken into account. Consequently, the test was reduced to the most efficient items and vocal and instrumental items were interpolated into the original body of piano test items. The final test of 50 items included examples developed from the songs found in the State Series as well as items developed in the improvisation classes of the Manhattanville Music Curriculum Program.

In the San Diego project validating classes had, for the most part, been given expanded musical experiences. They had participated in 15 Young Audience type programs and Saturday Morning Conservatory instruction. Teachers of the validating sample were asked to rate their pupils as to musicality using a five point scale, purposefully forcing a "normal" distribution. Test items were again chosen from a preliminary test battery that significantly differentiated between those students who had been rated in the upper 27 per cent of their classes by their teachers and those who were rated in the lowest 27 per cent of their classes.

The test, as postulated, was thus empirically validated in that the items were chosen on the basis of their effectiveness in differentiating between those students who were considered to be most successful in music and those who were assessed as being least successful. Moreover, a multiple correlation of this criterion with a weighted composite of 65 items as predictors by means of a stepwise regressional analysis was made. Using teachers' ratings as the criterion variable, this analysis, after 50 steps, yielded a multiple correlation coefficient of .532 significant

at the .05 level.

This test - the California Test of Aesthetic Judgments in Music - was the test used in the evaluation of the Manhattanville Music Curriculum Program. It was used as well in studying the value of such tests in the measurement of musicality in the elementary schools of California. It meets the ordinary criticism which teachers offer to evaluation, since the test items were, in a sense, selected by the teachers. Since the teachers defined "musicality" by describing which students are its highest achievers, it could be said that the test is "measuring what is being taught," from the teachers' point of view.

In the spring of 1968, the reconstructed test of aesthetic judgments in music was cross-validated by administering the test to pupils in grades three, four, five, and six in each of six schools in the Chula Vista district, as well as to the junior high school students in the Conservatory program. An analysis of the test results showed that the test was, for the most part, appropriate for use in these grades. Forty-four of the total test items correlated significantly with teachers' ratings of success in music as well as with the total test score.

Reliability coefficients ranged from .687 obtained for the third grade sample, to .83 obtained for the sixth grade sample. Validity of the test, expressed as a correlation between test scores and teachers' ratings of their pupils as to musicality, was computed to be .47 for the sixth grade students and .49 for the junior high school validating sample.

The range of item difficulty, i.e., the percent passing a given item, was found to extend from .13 to .94 with a mean index of difficulty established at .40. Although Cronbach suggests that the ideal test should have a .50 difficulty level, the test's difficulty level was not deemed to be unexpected in view of the wide range of grades participating in the program.

It is also significant that the test scores reveal an appropriate growth gradient in the elementary school. The mean scores and standard deviations for the elementary school and the Conservatory classes are as follows:

	Third	Fourth	Fifth	Sixth	Seventh	Eighth	Ninth
Mean	16.86	18.29	19.33	23.06	28.30	29.10	30.93
S.D.	4.39	4.58	5.96	4.90	3.72	4.88	4.86

The preliminary evaluation of growth in aesthetic sensitivity to music was begun in the first week of the fall semester, 1968. The California Test of Aesthetic Judgments in Music was administered to the third, fourth, fifth, and sixth grade classes in each of the six schools composing the experimental and control samples and to the Conservatory students. The final testing was completed in the first week of May, 1969. Table V shows the results of the testing.

TABLE V

Growth in Aesthetic Sensitivity
San Diego "Impetus for the Creative Arts"

School	Grade	N	Pre-test Mean	S.D.	Post-test Mean	S.D.	Signif. Level
Bo.	3rd	37	15.72	4.43	16.94	4.28	
	4th	31	17.15	4.63	19.38	4.43	.01
	5th	26	15.65	4.02	18.32	4.74	.01
	6th	28	18.14	3.67	22.92	3.29	.01
Hi.	3rd	35	15.85	5.16	16.34	4.97	
	4th	30	16.86	4.39	18.73	5.60	.05
	5th	29	18.29	4.58	18.66	5.07	
	6th	27	21.14	4.19	23.06	4.90	
Mu.	3rd	28	19.29	3.61	19.00	6.14	
	4th	34	18.08	5.54	22.53	5.30	.01
	5th	36	18.67	5.15	22.91	5.48	.01
	6th	27	22.20	5.04	24.50	4.35	
Fr.	3rd	40	12.29	6.49	15.51	4.62	.01
	4th	46	15.89	3.47	17.65	3.98	.01
	5th	36	17.05	4.59	16.61	4.02	
	6th	41	18.94	5.69	19.56	2.68	
Fl.	3rd	31	12.15	5.45	13.92	4.68	
	4th	45	13.92	4.24	17.65	4.02	.05
	5th	36	16.84	4.67	21.58	5.13	.01
	6th	33	24.60	4.60	25.81	4.60	
Ro.	3rd	32	17.66	4.50	19.94	4.85	.01
	4th	34	19.81	5.69	19.06	4.82	
	5th	21	17.33	5.26	19.51	4.52	.05
	6th	21	18.19	2.48	20.33	4.30	.01
Cons.	6th	6	20.76	5.62	26.52	4.67	.01
	8th	53	29.15	4.17	33.00	4.71	.01

CHAPTER III

RESULTS OF THE INVESTIGATION

Source of Data

Two evaluation programs were available to this researcher for obtaining the required data for testing the hypotheses: 1) a Program of Evaluation of Selected Test of Musical Achievement carried out by the California State Department of Education for possible statewide use, and 2) the evaluation of student achievement in the Manhattanville Music Curriculum Program.

I. Evaluating Selected Tests for Statewide Music Testing

Recent legislation in California, granting greater autonomy to local school boards in determining curriculum, was tempered to some extent by companion legislation making statewide testing mandatory in certain subject-content fields. Music was not excluded from this category.

The research to which this investigation allied itself was undertaken to determine the efficacy of available music achievement tests in measuring growth in musicality of children in the 4th, 6th, and 8th grades in public schools of California. It was undertaken under the aegis of the Northern California Music Educators' Association and the State Department of Education. Data from this research made it possible to compare the efficacy of the California Test of Aesthetic Judgments in measuring musical growth with six other tests commercially available. 5,982 subjects from 36 schools throughout California comprised the sample. The time of the experiment was the school year of 1969-1970.

II. Evaluation in the Manhattanville Program

The Manhattanville Curriculum was not developed as a methodology applicable at a particular grade level or for one type of class structure. It is, rather, intended to serve as an outline of educational activities based on principles which are apropos for all students at all levels of learning. It should be regarded more as

a way of learning than as a set of tactical plans which should be rigidly followed in the classroom. Only in the area of classroom procedure, which may be regarded as a style of operating, is any rigidity implied. In this area of educational style, adherence to three basic principles is considered most urgent. These include dedication to a learning process of personal discovery through creative exploration, avoidance of fragmentation in consideration of basic musical elements and organizational factors, and a concise delineation between concepts and skills.

The crux of the curriculum obviously lies in the use of musical improvisation by the pupils to discover musical concepts. Each work therefore must be "composed" for the forces at hand and each work is performed for analysis and criticism. Listening guides used by members of the program reveal a rather high sophistication of the analysis procedure. The same techniques of analysis are practiced by the students while listening to recordings intended to supplement and reinforce the concepts developed through improvisation. Undoubtedly this experience of critical analysis influenced perception and judgment-making required in the testing program.

Evaluating Selected Tests for Statewide Music Testing in California

Rationale

Central to the problem of evaluation in any field is the assumption of expectancies or standards. Evaluation tends to improve as the goals of the educational enterprise are made clear. Musicians are notably non-conformists, however, where the matter of fixed goals are suggested. Through empirical evidence they are quite cognizant of individual differences in musical aptitude and know full well the differing elementary school musical opportunities. Due, in part, to our highly mobile society and the differing emphases necessarily placed on the musical endeavors within California schools, music teachers wisely advocate the doctrine that the same music education is not necessarily the best education for all children.

The first limitation recognized for this evaluation project, therefore, lay in defining goals for music education which would receive some universal acceptance. The researchers proposed to rely on two sources in order to assure this consensus: (1) objectives of music education found in the state music texts, and (2) those expectancies enunciated in the proposed "Framework for Music Education in California." For example, the Framework believes the broad purpose of music education for all youth is to develop aesthetic sensitivity to music. These objectives gave direction to the selection of evaluation instruments suitable for this study.

The second limitation recognized in this endeavor arose through the selection of students who comprise the testing sample. It would be hoped that the sample indeed represents the population from which it was drawn as to socio-economic level, intelligence, musical opportunity, teacher competency, race, sex, etc.; therefore, considerable care was given to the choice of students to be evaluated. Moreover, the sample was extraordinarily large. It included 5,912 students from 36 schools participating in the testing program. A large geographical area was represented in the sample. Classes taught by regular classroom teachers were sought as well as those taught by music specialists. Urban and suburban areas were represented somewhat in the proportion in which they comprise the total state population. Extremes in socio-economic levels, intelligence, and musical opportunities were represented, though it was not feasible to control these factors prior to testing. Nevertheless, the results of this research can be generalized only to the extent to which the testing sample represents the total population.

The third limitation this research recognized is the paucity of standardized music achievement tests suitable for measuring achievement in the elementary school music classroom. It was germane to the purpose of the research that only those tests with published norms, machine scorable test forms, and which purport to measure musical achievement thought desirable by California music educators be considered. Standardized tests of musical achievement are unquestionably in short supply though there is a formidable array of tests described in doctoral dissertations, for which only the author's limited observations constitute the validating data.

Description of the Tests Studied

Colwell Music Achievement Tests

The first tests chosen for evaluation for possible state-wide testing in California were the Colwell Music Achievement Tests (M.A.T.) I and II.[1] The reason for their choice stems from the observation that the author's criteria for validation parallel that of this research. The rationale underlying the development of MAT is that music in the schools has content which is universally basic

---

[1]Richard Colwell, Music Achievement Tests I and II, Follett Educational Corporation, Chicago, 1969.

and which can be easily measured. This content is not a specific body of factual items since these items might differ widely in both area and quantity from school to school. Rather, it is a set of skills and understandings the pupil must have as he participates in making music or in listening to it.

The Colwell tests modestly do not claim to measure musicality except in a negative way. "The pupil who has failed to achieve in those auditory skills as covered by MAT is not positively musical or a musician."

The tests are intended to diagnose strengths and weaknesses for both pupils and teachers, help to mold objectives by implying certain standards of attainment, and to compare the outcomes of various methods of teaching. Their validity is declared to be the degree to which they measure the goals of music education as found in the several elementary music series. Thus, there is an apparent face validity which corresponds to that sought for this research. If the tests do reliably measure what they purport to measure, one need only to ascertain if what they measure is being taught or should be taught in California schools. To this end MAT scores were correlated with teachers' ratings of their students' musicality, that is, success in music from the California teachers' point of view.

Tests III and IV of MAT, recently standardized, were also made available for this evaluation. Since these tests are addressed to the stated objectives of junior high school music series, they were administered to 6th and 8th grades in this investigation.

Following are descriptions of the four Colwell Tests of Musical Achievement (MAT):

MAT I (68 Items)

Part I: Pitch Discrimination

Sub-test A of 15 items compares two tones. The testee decides which is higher or whether both tones are the same.

Sub-test B compares three tones. The testee must decide whether the first, second, or third is the lowest.

Part II: Interval Discrimination includes:

Sub-test A of 10 items comprised of three tone patterns in which the testee must decide if the pattern is scalewise or if the pattern leaps.

Sub-test B of 18 items in which the testee determines if the musical phrase is generally scalewise or if it generally leaps.

Part III: Meter Discrimination

This test is composed of 14 phrases each of which the testee must judge whether it moves in twos or in threes.

MAT II (78 Items)

Part I: Major and Minor Mode Discrimination includes the following:

Sub-test A in which the subject listens to two chords in each question and decides if they are major or minor chords.

Sub-test B - Phrases. The subject decides if each phrase is in major (M) or minor (m) or if it changes mode (c).

Part II: Feeling for Tonal Center is comprised of two sub-tests:

Sub-test A - Cadences. The subject listens to four chords and decides whether the first, second, third, or none of the succeeding three tones is the key tone.

Sub-test B - Phrases. After listening to each phrase the subject again is asked if either of the succeeding three tones is the key tone.

Part III: Auditory-Visual Discrimination also has two sub-tests:

Sub-test A - Pitch. The subject is asked to listen to four measures and to fill in the blank below every measure played differently in pitch from the notation. If the measure is correct, the box marked //0// is filled.

Sub-test B - Rhythm. The subject is asked to fill in the blank below every measure played differently in rhythm from the notation.

Colwell has recently produced two tests to extend the scope of MAT I and II. Standardization data are presently in press.

MAT III (74 Items)

Part I: Tonal Memory

Measures the ability of the testee to recognize changes in pitch within a four tone chord and the arpeggiated chord which follows.

Part II: Melody Recognition

Measures the ability to recognize in which part a melody is being played. The melody is first played on the piano and then repeated by a string trio with two harmonies added.

Part III: Pitch Recognition

Measures the ability to hear a tone in relation to the key tone. The subject decides which of three pitches matches the second written note. The first tone played is always the key tone.

Part IV: Instrument Recognition

Measures the subject's ability to recognize the sounds of various orchestral instruments first as solo instruments and secondly as "featured" instruments within the orchestra.

MAT IV (84 Items)

Part I: Musical Style

Sub-test A measures the subject's ability to distinguish different styles of music according to composer.

Sub-test B measures the subject's ability to distinguish music that is monophonic, homophonic, or polyphonic.

Part II: Auditory-Visual Discrimination

This test measures the subject's ability to compare rhythmic notation with music which is heard. The pitches are always written correctly.

Part III: Chord Recognition

Part III measures the subject's ability to recognize a chord the second time he hears it. One chord is played followed by three trial chords. The task is to select the trial chord which sounds like the original chord.

Part IV: Cadence Recognition

This test measures the testee's ability to recognize a full cadence, a half cadence, or a deceptive cadence as they occur in musical phrases.

Gordon Musical Aptitude Profile[2] Test S - Musical Sensitivity

The second test to be examined was chosen heedless of the advice of its author, Edwin Gordon, University of Iowa. Gordon questions the use of an aptitude test as a measure of musical achievement and doubts that public school music education is reflected by scores on the test. Relative stability of individual test score (in fact, a slight regression upon second testing) may indeed cause him to question the wisdom of its inclusion as an achievement test. There are at least two reasons for its consideration, however. First, the test results correlate quite well with teachers' estimates of success of their pupils. Secondly, there is a grade by grade growth gradient in the published norms. Here, as an example, are the mean scores for Test III (Musical Sensitivity):

---

[2]Edwin Gordon, Musical Aptitude Profile, Houghton Mifflin Company, Boston, 1965.

For Test S -- Musical Sensitivity

Grade	Mean	S.D.
4th	50.7	10.89
6th	56.0	11.39
8th	59.9	10.84

The following statements of the author, paraphrased from the test manual, further substantiated our decision to include the test in this survey:

> All aptitude tests are to some degree achievement tests. An aptitude test may be distinguished from an achievement test only to the extent that the generalized function of aptitude is relatively maximized and specifically taught course content material is relatively minimized. The Musical Aptitude Profile is designed to minimize factual knowledge so that the more basic factors of musical aptitude — musical expression, aural perception and kinesthetic music feeling — may be assessed. The tests are not concerned with historical or technical facts about music. Students are asked only to compare a selection with a musical answer and to decide if the selection and the musical answer are alike or different, exactly the same or different, or to decide which is indicative of a more musical performance.
>
> The Musical Sensitivity Test, which contains the preference measures, contributes to a more comprehensive appraisal of basic musicality, since musical creativity and expression are at least as important to success in music education as the abilities to perceive tonal and rhythmic relationship among notes in a musical phrase.

Gordon, of course, recognizes the factors of interest, opportunity for musical study, quality and quantity of formal music training, musical experience, opportunities to hear good music, parental encouragement of the study of music, cultural background, and physical coordination as being of considerable importance to success in musical endeavors. These factors may indeed be precisely the same factors ascribing the measurable results of music education.

The basic musical behavior measured by the Musical Aptitude Profile are classified into three main divisions: Tonal Imagery,

Rhythm Imagery, and Musical Sensitivity. Two separate sub-tests are provided for each of the non-preference tests, Tonal Imagery and Rhythm Imagery. They are Melody and Harmony for the former, and Tempo and Meter for the latter. The preference test, Musical Sensitivity, consists of three separate sub-tests. They are Phrasing, Balance, and Style. The tests consist of 90 original short selections composed for violin and cello by the author. Since the entire battery requires 3 class periods to administer, only Test III, Musical Sensitivity, was used in this research.

The preference tests require students to decide which of two renditions of a selection makes the better "musical sense." For each item in the Phrasing test, the same rendition is performed twice. The student must decide which rendition is performed with the better musical expression. In the Balance test, the second time the selection is performed it has a different ending. The student must decide which ending is better, both melodically and rhythmically for the selection. In the Style test, the same short selection is performed twice, but with different tempo. The second rendition is either consistently faster or slower than the first rendition. In all three tests the student simply makes the judgment for "1" or "2". If he has no preference he makes an "in doubt" response. In reality, the student responds to many elements of music in the items in the Phrasing, Balance and Style tests. In addition to the major factors of dynamics and tempo, the student is called upon to react to tone quality, intonation, and melodic and rhythmic contour.

Kyme, California Test of Aesthetic Judgments in Music[3]

The California Test of Aesthetic Judgments in Music is comprised of two sections: the first consists of piano items drawn from representative stylistic literature extending from the baroque era to the contemporary idiom. Short melodies are mutilated in terms of melody, harmony, rhythm and/or form. The student is asked if the paired renditions are the same or different, and if different which is more musically acceptable.

The second section of the test is comprised, in part, of songs chosen from the several school music series. Representative

---

[3]George H. Kyme, California Test of Aesthetic Judgments in Music, University of California, Berkeley, 1969.

instrumental pieces add further dimensions to the test by introducing the task of judging the appropriateness of musical style, tone quality, intonation and ensemble. The "correct" answers of the test are those answers given by the most musical students of the validating sample.

The test used in this evaluation was comprised of 40 piano items from the first test and 10 vocal and instrumental items from the second.

Gordon Iowa Tests of Musical Literacy[4]

The Iowa Tests of Musical Literacy by Edwin Gordon consist of measures of Tonal Concepts, Notational Understanding, and Rhythmic Concepts. There are six tests representing increasingly more difficult tasks but all measuring in the same categories. The Tonal Concepts test is divided into three sections: aural perception of major and minor mode, and reading recognition, that is, the detection of differences between the written and recorded versions of the items, and notational understanding. Levels I and II are recommended for 4th, 6th, and 8th grade subjects.

The Rhythmic Concepts test is likewise divided into three sub-tests. One measures aural perception of meter. The second evaluates reading recognition of rhythm, and the third measures the understanding of rhythmic notation.

CMP Musical Attitude Inventory[5]

> Competence is not enough — the educated person has feelings as well as skill. Competence is good, but it is not sufficient. All experience occurs in a feeling climate with an emotional overtone, in an affective state.*

---

[4]Gordon, Edwin, Iowa Tests of Music Literacy, University of Iowa Bureau of Educational Research and Service, (Iowa City, 1970).

[5]CMP Musical Attitude Inventory, unpublished tests from the Contemporary Music Project Office, (M.E.N.C., Washington, D.C., 1968).

*California Framework for Music K-12, loc. cit.

The purpose of administering the CMP inventory was to measure students' attitude towards the whole musical enterprise. Music education which proposes to develop aesthetic sensitivity must begin first of all with an open mind towards many kinds of music, not excluding contemporary music, the authors believe.

Methodology

Time, Place, Test Sample and Data Collected

In October, 1969, graduate students from the University of California, Berkeley, aided by music teachers in the field, administered the chosen tests to children in the 4th, 6th, and 8th grades of 36 schools in California. The number of students taking each test were as follows:

TABLE VI

Tests and Subjects in California Evaluation

Test	# Students Taking Test
Colwell, Music Achievement Test (MAT) I	1476
MAT II	1063
MAT III	489
MAT IV	431
Kyme, California Test of Aesthetic Judgments in Music	1496
Gordon, Musical Aptitude Profile Test S -- (MAP) Musical Sensitivity	310
Gordon, Iowa Tests of Musical Literacy	647
CMP Musical Attitude Inventory	312

In many instances the subjects were given two or more of the tests and teachers' estimates of the pupils' musicality were obtained. In the spring of the year the tests were again administered to 1000 children at each grade level. Certain tests were also administered to an independent sample, one which received music instruction but which had not been given the pre-instruction tests. This latter testing was instituted to determine the effect of practice in taking this test upon gains in test scores.

Statistical Treatment of the Data

The first step in analyzing the results was to ascertain the reliability and validity of each test for each grade. Once reliability and validity of the several tests had been calculated, several additional statistical processes presented themselves. They included the following:

1. Tests of the significance of the difference between mean scores of the three grade levels. It is held that significant gains from one grade level to the next indicate growth in music achievement and that the tests are sensitive to this change.

2. Correlation between the several tests. Intercorrelations tend to give credence to the validity of each measure but, more importantly, show the extent of overlapping and redundancy. Tests measuring precisely the same behavior as another do not contribute to information already obtained.

3. Analysis of covariance. This statistic was used to statistically equate dissimilar groups and compare post-test results to ascertain the differences in musical achievement of students taught music by specialist teachers and those taught by classroom teachers.

4. Finally, an item analysis was made to identify the items and sub-tests which tended to differentiate between the low achiever and the high achiever.

The first information sought in evaluation of the efficacy of each test was the reliability of the test. Test reliability reported here is the proportion of true variance accounted for by the examinees. It is obtained by regarding the matrix of item scores as a two-way design for analysis of variance without replication. The formula then is:

$r = \frac{Ve - Vr}{Ve}$ where Ve represents the variance for examinees, and Vr represents variance for the remainder sum of squares. The result is identical to that obtained from the Kuder-Richards Formula 20 for testing reliability. Table VII shows the reliability of each of the tests and the size of sample from which it was obtained.

From this table it may be said that all tests are sufficiently reliable for group testing for the grades recorded with the exception of the CMP Musical Interest Inventory.

TABLE VII

Reliability of the Tests Investigated

Test	Number	Grade	Reliability
MAT I	221	6th	.929
	223	8th	.909
MAT II	158	4th	.801
	78	6th	.850
	72	8th	.823
MAT III	112	8th	.840
MAT IV	80	8th	.724
Gordon-Iowa Test of Musical Literacy I	85	6th	.818
Gordon-Iowa Test of Musical Literacy II	85	8th	.877
Calif. Aesthetic Judgments	678	6th	.809
	223	8th	.845
CMP Musical Interest Inventory	140	6th	.529
Gordon-Musical Aptitude Profile	73	8th	.850

Validity of the Tests

Validity is often defined as how well a test measures what it purports to measure. An alternate definition, enlarging upon the first, defines validity as how well we know what a test measures -- regardless of what it purports to measure. The importance of item validity as a part of test validity leads to careful consideration of the kind of criterion measure which is valid itself as well as reliable with which the validity of each item or test may be evaluated. It is obvious that any validational problem centers around the characteristic of the criterion measure itself. Recognizing full well the dangers inherent in the process, it nevertheless seemed prudent for this experiment to use the teachers' estimate of their students'

musicality as the primary validation criterion. This decision was influenced by two considerations: (1) music teachers should have confidence that tests proposed for state-wide testing validity measure what they are striving to teach; and (2) it must be admitted, that for estimating the musical achievement of students, no other single measure of success in music equals the estimates of teachers who have observed their students in many musical situations. The validity reported in Table VIII, therefore, represents the correlation between test scores and teachers' ratings of their pupils. The ratings were made upon a seven point scale and a "normal" distribution was required for each class. The validity coefficients were computed for each grade level.

TABLE VIII

Validity of the Tests

Test	Grade	Number	Validity*
MAT I	4	249	.357
	6	312	.417
	8	323	.520
MAT II	4	294	.593
	6	417	.580
	8	324	.579
MAT III	6	72	.280
	8	58	.423
MAT IV	6	72	.260
	8	62	.652
Calif. Test Aesthetic Judgments	4	80	.335
	6	678	.487
	8	223	.533
MAP	6	72	.462
	8	61	.531
Iowa Musical Literacy I	6	72	.25
Iowa Musical Literacy II	8	61	-.10

*Correlation with teacher ratings.

TABLE IX

Summary of Means, Standard Deviations, t Tests and Intercorrelations (Fall Testing)

Test	Grade	Sample Size	Means	S.D.	t Test	Significance Level	Reliability	Validity	Correlation Matrix MAT	MAP	Calif.	Iowa I
MAT I	4	249	33.80	13.67			.84	.35	II.16	.26	.19	
	6	312	35.64	13.30	1.61	not sign.	.88	.41	.72			
	8	323	40.06	10.57	4.64	.01	.90	.52	.66			.37
MAT II	4	294	26.11	8.29			.80	.59		.18	.26	
	6	417	33.17	10.85	6.72	.01	.85	.58		.44	.50	.46
	8	324	33.87	12.13	.49	not sign.	.82	.57			.34	
MAT III	4											
	6	58	31.41	11.36	2.36	.05	.84	.42	IV.26			
	8	61	36.36	9.07			.84	.32	IV.42	.18	.14	.34
MAT IV	6	58	35.11	8.69	6.94	.01		.26	III.26		.34	
	8	61	46.64	7.73			.72	.58	III.42	.44	.47	.14
MAP	4	35	53.08	12.71					I.26			
(Sensitivity)	6	38	57.36	10.05			.85	.53	II.18			.07
	8	61	60.13	10.42	.92	not sign.			IV.44		.19	
Calif. Aesth.	4	592	17.01	5.60			.51	.33	II.26			
Judgments	6	678	20.13	4.86		.01	.81	.48	II.50			
	8	223	26.76	5.06	5.56	.01	.84	.53	IV.47	.19		.22
Iowa Test of Music Literacy												
Level I	4	49	70.14	19.60					I.37			
Level I	6	94	76.37	22.55				.24	II.46	.07		
Level I	8	72	77.52	23.46	1.64	not sign.	.81		III.34			
Level II	6	50	66.20	13.95				.25				
Level II	8	61	67.21	21.52	.30	not sign.	.80	.38	IV.14		.22	
CMP Attitude	4	58	17.80	6.15			.529					
Inventory	6	140	21.01	6.40	2.87	.01						

Table IV presents some useful information:

1. The t test column shows which tests reflect significant growth from grade to grade.
2. The correlation matrix shows the intercorrelations among the various tests.

TABLE X

Evidence of Growth in Musicality (Fall and Spring Testing)

Test	Grade	No. Cases	Sept.	S.D.	Feb.	S.D.	t-Test	Significance Level
Colwell: MAT I-II combined	8th	N = 50						
Pitch			13.66	4.65	17.41	5.33	2.61	.05
Interval			13.44	4.00	17.38	6.76	1.36	n.s.
Meter			13.80	5.21	18.45	7.61	2.42	.05
Aud/Vis - Pitch & Rhy.			41.04	5.50	44.83	8.91	1.73	.10
Major-Minor			6.57	1.91	8.22	3.03	2.21	.05
Tonal Center			3.33	1.46	4.96	2.30	2.87	.01
Total			93.57	38.08	111.29	32.48	.05	n.s.
Colwell: MAT IV	8th	N = 121						
Musical Style			5.72	2.24	8.10	2.23	5.39	.01
Texture			8.84	3.30	12.08	3.37	4.93	.01
Aud/Vis - Discrim.			4.44	2.38	5.62	2.72	2.40	.05
Chord Recognition			8.90	3.50	12.10	2.29	5.09	.01
Cadence Recognition			7.20	2.74	8.72	2.18	2.99	.01
Total			35.11	8.69	46.64	7.73	6.94	.01
Gordon: Iowa Tests of Musical Literacy II	8th	N = 61	67.21	21.52	79.11	23.30	4.32	.01
Gordon: MAP - Test S								
Part I	8th	N = 73	20.36	3.71	20.29	3.93		n.s.
Part II			21.61	4.04	21.52	4.24		n.s.
Part III			18.95	4.33	18.50	3.94		n.s.
Total			60.84	10.22	60.13	10.42		n.s.
Kyme: California Test of Aesthetic Judgments	4th	N = 350	16.70	5.39	16.76	5.30	.13	n.s.
	6th	N = 146	17.38	5.04	21.17	4.34	1.83	.10
(8th grade Instrumental Musicians)	8th	N = 53	29.15	4.17	33.00	4.71	8.34	.01

These data for MAT I, II, and IV were obtained from the same junior high school but not from the same students While there is observable growth in this sample's mean scores between the Fall and Spring testing for MAT I and II, the differences are not significant. The MAT IV testings did yield significantly different mean scores for each of the sub-tests as well as for the total score.

Where possible, intercorrelations coefficients were obtained between the several tests. This statistic serves to ascertain that the tests purporting to measure musical achievement are indeed measuring somewhat the same attributes but, conversely, assures the researcher that each test is somewhat unique in its measurement. Such information would indeed be helpful if the teacher proposed to administer tests comprised of the best sub-test from each of the tests. It should be remembered in interpreting these data that the validating sample was not the same for all testing. Moreover, it is impossible to equate the abilities of teachers to make estimates of their students' musicality, though an attempt was made to assure a "bell curve" distribution of the ratings made by each teacher.

In one instance, however, six schools from the same school district administered the California Test of Aesthetic Judgments in Music. Three of these schools received special musical instruction, the others were taught by the regular classroom teachers. In analyzing the results, an analysis of covariance statistic was used which, by considering the regression line, equates the two groups on pre-test scores and yields an adjusted post-test mean score as though the groups were indeed equal. Table XI shows the results by grades.

TABLE XI

Growth in Aesthetic Sensitivity - Music
Specialist Versus Classroom Teacher

Grade	Type of Instruction	Pre-test Mean	Post-test Mean	Adjusted Mean	Degrees of Freedom	F Value	Level of Sig.
4	Specialist	18.50	19.03	18.81	288	2.80	NS
	Classroom Tch	17.03	17.56	17.88			
6	Specialist	20.82	22.00	22.23	255	6.60	.01
	Classroom Tch	21.73	21.13	20.91			
8	Specialist	26.89	28.03	27.89	622	11.78	.01
	Classroom Tch	25.25	24.51	25.61			

Music educators often express concern over the tastes exhibited by teen-agers, as evidenced by their radio-listening habits. While they realize that tastes are imposed by disc jockeys and commercial interests not concerned with the education of

children, the question from an educator's view is this--does exposure to a limited style of music encourage musical sensitivity as it applies to "serious" music? To answer this question, a questionnaire was administered to certain students taking the California Aesthetic Judgments Test. The survey purported to determine the value which students place upon 50 selected musical activities. A comparison was then made of the responses of the quartile which scored highest on the Aesthetic Judgments Test and that quartile which scored the lowest. A second analysis was made between the total sample that had participated in the project's expanded experiences and the control classes which had not. From an analysis of the questionnaire these observations are presented in the following charts.

The measure undertaken in this testing program deserves mention because it, too represents an attempt to assess the results of public school music to bring about a consensus in the affective domain. This type of measurement, of course, is important though usually unreliable. The instrument of evaluation used was the CMP Musical Interest and Attitude Inventory. Although there are no "right" and "wrong" answers in the measurement of preferences and attitudes towards music, these data were submitted to statistical treatment. The data presented on the following page as a chart clearly shows the similarities and differences in the preferences of a sixth grade sample which was composed of those students who scored highest and lowest on the Test of Aesthetic Judgments. The two contrasted samples are found to be much more alike than different. A similar chart illustrating the interests and attitudes of high and low socio-economic schools produced almost identical results. From the charts it may be concluded that the schools run a poor second to the pressures of "peer-approval" in developing children's attitudes towards music.

It is suspected that this inventory will not serve music education's needs in evaluation. It simply fails to show the effects of classroom instruction. However, each teacher may wish to evaluate his offerings in terms of the results of such a survey, since listening to many kinds of music and performance mark the most musical student from the least musical.

The foregoing analysis is of interest because it describes the kind of person who is likely to score highest on the California Test of Aesthetic Judgments in Music. Equally important is the comparison of the variable characteristics typifying those students who had participated in the PACE Programs. The following chart,

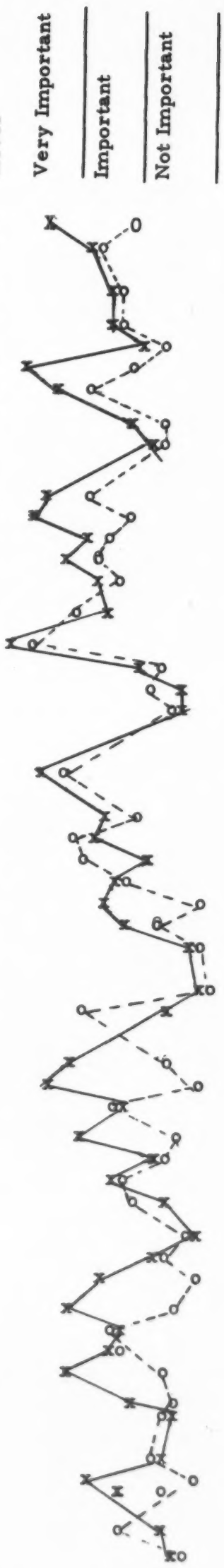
Comparison of Attitude Scores of High and Low Achievers on California Test of Aesthetic Judgments

How important to you are these experiences in music?

X = High Achievers

O = Low Achievers

	Significant Level
1. Listening to many kinds of music	.01
2. Listening to the same things over again	...
3. Responses associated with listening	
(a) Rhythmic movements	...
(b) Singing	...
(c) Beating time	...
4. Learning an instrument	.01
5. Learning to sing	...
6. Learning to read music	...
7. Memorizing music	...
8. Performance (playing an instrument or singing)	
(a) For yourself	.05
(b) For others	.01
(c) In a small group	...
(d) In a large group	...
(e) In other activities (theater, film, dance)	...
9. Classroom instruction	...
10. Private instruction	...
11. Group instruction	...
12. Reading historical materials (about music)	...
13. Reading biographical materials (about composers or performers)	...
14. Listening to the radio	...
15. Listening to recorded music	
(a) Chamber music	...
(b) Rock and Roll	...
(c) Jazz	.05
(d) Religious	
(e) Symphonic	.01
(f) Vocal	...
(g) Country Western	...
(h) Electronic sound	...
16. Listening/viewing televised music programs	.01
17. Attending live performances	
(a) To hear a specific work or composition	.01
(b) To hear a soloist	.01
(c) To hear a chamber group	...
(d) To hear an orchestra or band	.01
(e) To hear a dramatic work	...
(f) To hear religious music	...
(g) To hear jazz	...
18. Experimenting with musical instruments	...
19. Experimenting with sound-producing devices	...
20. Writing music	.01
21. Taking part in music competitions	.01
22. Discussing music with your fellow students	...
23. Meeting noted musicians	.01
24. Knowing different kinds of music	
(a) Chamber music	.01
(b) Rock and Roll	...
(c) Jazz	...
(d) Religious	...
(e) Symphonic	.01
(f) Vocal	.01
(g) Country Western	...
(h) Electronic sound	...



derived from the questionnaire, shows the mean score responses on a three point scale of the "X" group of participants in the program and the "O" sample which did not. The significance levels of the differences of the means is also indicated.

It is evident from the chart that the students of the experimental and control schools are more alike in their values than they are different. In only 28% of the musical experiences was there a difference significant at the .05 level. The important consideration in analyzing the data, however, should be in which values the two groups differ.

The evidence indicates that the experimental schools put a significantly higher value upon the following musical experiences:

	t ratio
1. Listening to many kinds of music	2.87
2. Listening to the same music over again	3.29
3. Singing	4.53
4. Learning to read music	2.53
5. Memorizing music	3.15
6. Listening to chamber music	2.60
7. Listening to religious music	3.11
8. Listening to vocal music	2.55
9. Attending live performances to hear soloists	2.21
10. Attending concerts of religious music	3.26
11. Writing music	2.23
12. Knowing religious music	2.81

The control classes, on the other hand, differed significantly from the experimental in the high value they placed upon the following experiences.

1. Listening to the radio	2.26
2. Listening to electronic music	3.41

Comparison of Experimental and Control Schools on Responses to Questionnaire

How important to you are these experiences in music?

X = Experimental

O = Control

	Significant Level
1. Listening to many kinds of music	.01
2. Listening to the same things over again	.01
3. Responses associated with listening	
(a) Rhythmic movements	...
(b) Singing	.01
(c) Beating time	...
4. Learning an instrument	...
5. Learning to sing	.05
6. Learning to read music	...
7. Memorizing music	.01
8. Performance (playing an instrument or singing)	
(a) For yourself	...
(b) For others	...
(c) In a small group	...
(d) In a large group	...
(e) In other activities (theater, film, dance)	...
9. Classroom instruction	...
10. Private instruction	...
11. Group instruction	...
12. Reading historical materials (about music)	.10
13. Reading biographical materials (about composers or performers)	...
14. Listening to the radio	.05
15. Listening to recorded music	
(a) Chamber music	.05
(b) Rock and Roll	...
(c) Jazz	.10
(d) Religious	.01
(e) Symphonic	...
(f) Vocal	.05
(g) Country Western	...
(h) Electronic sound	.01
16. Listening/viewing televised music programs	...
17. Attending live performances	
(a) To hear a specific work or composition	...
(b) To hear a soloist	.05
(c) To hear a chamber group	...
(d) To hear an orchestra or band	...
(e) To hear a dramatic work	...
(f) To hear religious music	.01
(g) To hear jazz	...
18. Experimenting with musical instruments	...
19. Experimenting with sound-producing devices	...
20. Writing music	.05
21. Taking part in music competitions	...
22. Discussing music with your fellow students	...
23. Meeting noted musicians	...
24. Knowing different kinds of music	
(a) Chamber music	.10
(b) Rock and Roll	...
(c) Jazz	...
(d) Religious	.102
(e) Symphonic	...
(f) Vocal	...
(g) Country Western	...
(h) Electronic sound	...

Measurement of Aesthetic Sensitivity in the Manhattanville Music Curriculum Program

The customary research design for measuring growth in any curricular innovation is to measure--teach--and measure again. The difference between pre-instruction and post-instruction test results then is assumed to be due to the learning experience intervening. Such an assumption, of course, disregards the many factors inherent in the testing procedure, such as the effect of practice in taking the test, the differing testing conditions (the passing of an airplane, a dropped book), the effect of fatigue and particularly the boredom of taking the same kind of test over again. There is considerable evidence that young children tend to score lower on a second administration of the California Test of Aesthetic Judgments in Music than on the initial testing, particularly if little or no musical instruction has been presented between testings.

In evaluating and comparing the relative achievement of students in the Manhattanville program, no assumption of equality of prior musical experiences should be made. Students drawn from the wide stretches of the country, from Oregon, Texas, Colorado, New York, New Jersey, and Connecticut, represent the widest spectrum of socioeconomic subcultures. Any application of standards or norms which might be appropriate for all of these schools would be patently ridiculous. Effective controls to simulate equality between samples could not be instituted in the public school situations. However, two strategies were instituted to hold two important factors constant, and to weigh an extraneous element contributing to differences in test scores. They were 1) to utilize only the classes taught by the music specialist in each school, one of which would not receive the experimental instruction; 2) to randomly select classes in each grade to which the post instruction alone would be administered in order to ascertain the effect of repeated testing upon final test results.

TABLE XII

Results of the Testing
Manhattanville Music Curriculum Program

School	Grade	N=	Pre Test	S. D.	Post Test	S. D.	t Test	Signif-icance
				FOURTH GRADE				
Zero Control	4th	18	17.72	4.05	17.52	5.73	.22	N. S.
	4th	21	19.32	4.61	19.32	4.51	.00	N. S.
	4th	28	17.66	4.52	18.02	4.33	.49	N. S.
	4th	17	16.52	4.51	17.62	3.05	.91	N. S.
	4th	23	--	--	20.72	4.38		
				FIFTH GRADE				
Zero Control	5th	35	19.30	4.68	19.12	4.25	.29	N. S.
	5th	42	--	--	23.23	4.85		
	5th	34	--	--	23.60	5.32		
	5th	46	19.00	4.28	20.17	5.66	1.48	N. S.
	5th	34	--	--	17.42	4.50		
				SIXTH GRADE				
Zero Control	6th	24	22.35	5.48	20.20	5.79	1.73	N. S.
	6th	23	20.70	4.28	21.60	4.08	1.14	N. S.
	6th	21	--	--	24.10	5.73		
	6th	16	--	--	23.12	4.06		
	6th	21	--	--	22.60	7.55		
				SEVENTH GRADE				
Zero Control	7th	56	22.75	4.08	22.85	4.68	.14	N. S.
	7th	40	22.30	4.58	23.25	4.67	1.83	N. S.
				EIGHTH GRADE				
Y	8th	15	24.93	6.07	26.73	4.18	1.77	N. S.
	8th	16	22.00	5.63	26.25	4.08	2.73	.05
	8th	21	23.33	4.18	24.88	4.41	1.24	N. S.
	8th	15	21.50	4.81	26.71	3.91	5.32	.01
	8th	26	26.31	5.36	24.50	4.66	.80	N. S.
	8th	72	24.00	5.12	29.32	4.89	14. 70	.01
Zero Control	8th	91	22.96	4.78	23.31	8.29	.24	N. S.
Exp Control	8th	30	19.40	3.48	23.46	3.75	2.76	.05
Instr Control	8th	48	25.95	4.83	29.08	3.57	3.02	.01
				TENTH GRADE				
Y	10th	32	28.75	4.29	31.12	3.77	1.89	.10
T	10th	18	28.61	3.77	34.30	7.45		.01
Control	10th	19	--	--	28.20	6.23		
Control Chr	10th	38	--	--	20.72	4.81		
Control Orch	10th	117	--	--	21.05	4.64		

TABLE XIII

Comparison of Aesthetic Judgment Scores of Students Participating in California PACE Projects with Those of the Manhattanville Program (Post-Instructional Testing, California Test of Aesthetic Judgments in Music)

Grade	Published Norms (18 states)	Project ORFF Madera	Impetus for Creative Arts Chula Vista	Kodaly-Richards Portola Valley	Manhattanville Curriculum Project
4th	n=812 m=17. 32 SD=5. 01	n=350 m=16. 76 SD=5. 30	n=242 m=17. 26 SD=4. 70	n=125 m=18. 50 SD=5. 26	n=84 m=18. 06 SD=4. 52
6th	n=1003 m=21. 05 SD=5. 04	n=300 m=18. 54 SD=5. 39	n=258 m=21. 73 SD=4. 34	n=131 m=25. 90 SD=4. 99	n=122 m=22. 35 SD=5. 48
8th	n=1145 m=25. 88 SD=5. 50	n=78 m=21. 12 SD=6. 04	n=202 m=24. 53 SD=5. 78	n=12 m=29. 52 SD=2. 78	n=135 m=28. 25 SD=4. 67
8th grade instru-mental-ists	n=645 m=28. 80 SD=5. 12		n=52 m=32. 60 SD=4. 48		n=26 m=31. 12 SD=4. 66

Table XIV shows the results of the experimental music programs as measured by the California Test of Aesthetic Judgments in Music.

In a study of the nature of growth in aesthetic sensitivity to music in the schools, how children differ in performing judgmental tasks is perhaps more enlightening than mere knowledge of the extent of this difference. Consequently, the program for machine scoring of the answer sheets provided for four subscores in addition to the total score for each subject. The subscores were categorized as melody items, harmony items, rhythm items and items concerned with musical form. A fifth category, items dealing with "performance style," comprises the remainder. Ten items in each of the five

categories constitute the 50 item test. Table XIV shows the change in percent of correct responses in each category of children were tested in grades 4, 6, and 8. The subjects were selected from a typical community of each of the experimental programs, with the exception of the San Diego Saturday Morning Conservatory Program which drew from several city schools.

TABLE XIV

Percent Passing Items in Each Category

School	Grade	N=	Melody	Harmony	Rhythm	Form	Mus. Style	Total Score
Madera	4th	103	32	40	48	32		16.84
ORFF	6th	20	56	56	56	81		24.90
Project	8th	150	56	68	79	50		25.58
Portola	4th	125	43	48	55	37		18.50
Valley	6th	131	59	70	76	52		25.92
Kodaly Curric.	8th	12	62	79	93	60		29.52
Chula	4th	190	42	47	61	47	42	18.08
Vista	6th	193	48	56	63	56	45	22.20
Impetus for the Arts S. D. Conservatory	8th	52	71	87	95	87	61	32.60
Manhattan-	4th	23	49	51	50	20	23	19.51
ville	6th	17	58	68	56	24	30	23.82
Curric. Program	8th	18	67	85	65	36	31	28.61

The program for machine scoring of the test blanks also permitted another interesting observation to be made, a possible clue as to the source of variance between scores of the Manhattanville and California samples. It will be recalled that the test directions provided for a response in one of three columns:

1. The first performance is better than the second
2. The second performance is better than the first
3. The two performances are exactly alike.

Actually, no performances were duplicated! The third choice was included as a foil and tended to increase the reliability of the test by minimizing the advantage of guessing.

If test scores were the product of chance alone, it might be assumed that one third of all responses would fall in the third or same column. Inspection of the responses marked in column three indeed showed that approximately one third of all responses were in that column for grades up to and including grade four. This observation held true for California and Manhattanville subjects alike. In the eighth grade, on the other hand, the third column was almost totally ignored by the Manhattanville students while the California students still responded eleven times out of fifty, i.e., 22 percent of the time in this manner. The implication is obvious: scores obtained through chance alone should be higher if only two categories are possible than when three are utilized. Since the Manhattanville scores are not higher, one might conclude that while these students are more perceptive, that is, they do detect the small differences in performances which make them not alike, they do not as readily endorse the expected preference. This evidence says something of the ambiguity in taste, the open-mindedness toward the unexpected, the prevention of closure sought by the Manhattanville improvisational experiences. It is difficult to explain the categories in which the Manhattanville students do not excel, however. As may be seen in Table XIV, only in items of rhythm and form are they significantly different from their California counterparts.

CHAPTER IV

DISCUSSION

The interpretation and conclusions will be organized under two headings: 1) the value of aesthetic judgments in the assessment of musical achievement, and 2) an evaluation of the Manhattanville Music Curriculum Program.

The Value of Aesthetic Judgments in the Assessment of Musical Achievement

The questions central to this investigation were presented in the form of hypotheses in the introductory chapter. They were stated not modestly as null hypotheses, but rather as alternative or opposing hypotheses to be tested. In order to assess the effectiveness of this research it would seem prudent here to review those questions and marshall the relevant evidence toward each question in turn.

It was postulated that making aesthetic judgments, described as a measurable manifestation of conceptual development, may serve as the organizing factor of the elements of auditory imagery, and that, therefore, students who score highest on tests of aural perception and on tests of musical concepts ought also to score highest on tests of aesthetic judgments in music. Table IX, showing the intercorrelation between the several types of tests used in the California Statewide Music Testing Program, do indeed indicate that high achievers in perception tasks score highest on the tests of aesthetic judgments. Positive correlation coefficients with the Colwell Music Achievement Tests I, II, III and IV, with Gordon's Musical Aptitude Profile, and with the Iowa Tests of Music Literacy, ranging from .19 to .50 give strong evidence as to the truth of this hypothesis.

The first hypothesis also declared that correlations with teachers' ratings should be higher with aesthetic judgment test

scores than with simple perception test scores. From Table XIV it may be seen that those correlation coefficients expressed as validities of the several tests show that two tests of judgment, the Gordon MAP and the California Test of Aesthetic Judgments in Music do exceed all others at the sixth grade level with the exception of MAT II and they compare favorably with MAT II and IV at the eighth grade level.

The contributions of aesthetic judgments in the assessment of achievement in music was determined as a product of the computer program for machine-scoring the Aesthetic Judgments Test. In the scoring process, tests were first scored as simple perception tests in which students were credited with the recognition that the two performances comprising each item were the same or different. The second run of the program then credited the preferences expressed for either performance of items identified as being different. Correlations with teachers' ratings of their pupils as to musicality increased from .164 to .896 as a result of the addition of the preference component. This extremely high correlation obtained from the eighth grade classes of one of the schools in the Manhattanville Project was not commonplace. Nevertheless, it substantiates the postulate that in the aesthetic experience, perception alone is not enough. Reacting or responding is also required. Aesthetic judgments may indeed serve as the quasi-organizing factor of the elements of auditory imagery.

The second question asked, in the form of an hypothesis, was whether or not aesthetic judgments are culturally derived and therefore subject to the educational process. It was postulated that there should be a growth from grade to grade in mean scores on the test of aesthetic judgments which, in turn, would be significantly affected by the quality of the music programs in the schools.

The evidence is perfectly clear that scores on the California Test of Aesthetic Judgments in Music are affected by musical instruction in the schools. The following table from the Chula Vista PACE project will serve to illustrate the growth gradient found in the elementary schools.

	Third	Fourth	Fifth	Sixth	Seventh	Eighth	Ninth
Mean	16.86	18.29	19.33	23.06	28.30	29.10	30.93
S. D.	4.39	4.58	5.96	4.96	3.72	4.88	4.86

To say, however, that this growth is significantly affected by the quality of the music programs, one must make some

assumptions as to what constitutes quality of instruction. Here, two possible assumptions were made which are admittedly open to question yet have sufficient merit to warrant consideration:

1) An experimental program, taught by music specialists is indeed of higher quality than control classes taught by regular classroom teachers.

2) The quality of musical experience in the second year of experimental programs ought to be improved by virtue of the experience gained and the opportunity afforded to reject inefficient techniques. Using these criteria for "quality" definition in elementary school music, the evidence is not entirely clear.

A comparison of each of the major experimental programs found in Table XIII, with "norms" established upon a random sample drawn from eighteen states, shows that in most of the experimental classes this postulate is true. A comparison of test results obtained in California comparing the effectiveness of classroom teacher and music specialist (found in Table XI) shows that growth in aesthetic sensitivity is statistically superior in classes taught by music specialists in the sixth and eighth grades.

A comparison made of scores obtained at the end of the first year of a PACE project and those obtained at the end of the second shows that higher gains may be anticipated at the end of the second year only in the upper grades. The following chart showing test scores in two consecutive years of the Madera PACE project illustrates this phenomenon.

TABLE XV

Comparison of First and Second Year Scores for Primary Grade Students in the Madera "Orff" Project

When Tested	Grade	N=	Pre	Post	Adj. Mean	S. E.	F Value	Signif.
1967	2nd	289	17.93	17.30	17.15	.52	.91	N.S.
1968	3rd		16.08	17.93	17.87	.52		
1967	3rd	413	16.52	17.49	17.41	.43	4.72	.01
1968	4th		15.59	18.67	18.75	.43		

The third hypothesis is concerned with the nature of the ability to make aesthetic judgments in music. The hypothesis, described as an opposing one to the second, simply asks this question, "Does the ability to make judgments of appropriateness primarily reflect innate aesthetic tendencies which formal instruction in the schools can influence little if at all?" If so, then it should be expected that individual differences within any single class or grade would be significantly greater than mean differences between any two consecutive grades.

Table VIII shows that while standard deviations of the means of each grade do not vary widely from sample to sample they are indeed greater than the differences in mean scores between grades. It is obvious that there are individual differences in the ability to make aesthetic judgments which for lack of a better description is often termed "inherent musical aptitude." The recognition of individual differences, however, does not denigrate the contribution of school music in developing this ability to make aesthetic judgments. The failure to gain in the ability to make appropriate judgments by the zero-control groups, irrespective of the program of which they are a part, indicates that talent alone cannot account for increased sensitivity. One of the interesting postulates in aesthetic education is that tastes are indeed cultural derivatives springing from innate musical tendencies.

Evaluation of the Manhattanville Music Curriculum Program

Tables XII and XIII show the mean scores for the several grades measured in the Manhattanville Music Curriculum Program on the California Test of Aesthetic Judgments in Music. Also included are comparable grades from three PACE projects in music from California. It must be emphasized that no justifiable comparisons can be made between the samples. No attempt was made to equate the participants, no controls placed upon the frequency or length of the class periods, and no consideration given to the expertise of the teachers involved in carrying out the programs. The table does show a growth from grade to grade within each program from which certain inferences may be made. It was possible to institute some controls within each school of the several programs. Consequently, the evidence presented within each school is more enlightening than evidence gathered from total populations. In a word, the unique contribution of the Manhattanville program needs to be assessed in terms of its own expressed goals, and,

particularly in terms of growth in aesthetic sensitivity within each grade and school.

First of the expressed purposes of the program was to develop aesthetic sensitivity to music. Table XII indicates the significance of the difference between pre and post testing scores. It will be seen that only at the eighth and tenth grades is the measured growth in aesthetic sensitivity statistically significant for the majority of the classes. It will be noted, also, that children taking the test only as a post instruction measure scored higher than those taking the test for the second time.

The second of the expressed goals of the Manhattanville Program was "to prevent closure." That is to say, one purpose of the program was to foster a tolerance for the unknown as well as an appreciation for our musical heritage. In this endeavor the Manhattanville program was more successful. The evidence comes rather indirectly as a result of an analysis of variance. Item analyses showed, in the first place, that the most musical students in the program tended to answer in the same way that their California counterparts responded. Conformity, however, expressed as the percent passing any item drawn from the several musical periods, Baroque, Classical, Romantic, Impressionistic, or Modern, varied widely within a single grade depending upon the community or program supplying the data.

Yet students from the Manhattanville Program were as much like their California counterparts as they were like each other. Correlations with teachers' ratings of their pupils' musicality were often higher for item scores on the Manhattanville program than within the California samples. Students of both populations were equally inclined to prefer the altered harmonizations of the contemporary, bi-tonal examples of Milhaud, Hindemith and Ravel. Their predilection for a diatonic harmonization of these pieces reflected an adjustment to this harmony of our ubiquitous folk singers. The Manhattanville students held the same values as the California students but held them more intensely. It must be admitted that from the item analyses, gains made from pre to post instruction testing might be accounted for in terms of increased perception of the minute differences in the two performances comprising each item rather than from any standardization of taste resulting from the instructional program. The observation that the Manhattanville students tended to avoid the testing foil--"the two pieces are exactly the same," ordinarily would mean that by limiting choices to two of the three possible categories, the subject would

score half of the items correctly, purely through guessing. Since there was no significant superiority of those in the Manhattanville program over the California students who still were inclined to answer more frequently in the "Same column," it must be concluded that a real ambiguity on the part of the Manhattanville student as to the "appropriate" usage of the elements of music caused him to respond equivocally to the two performances comprising each item. This tolerance for the strange which Thomas described as "prevention of closure" would suggest that the Manhattanville program has indeed opened minds to contemporary sounds. A judgment as to the desirability of this outcome is not in the province of this research, inasmuch as a declared value of aesthetic judgment testing is rather to know oneself than to standardize taste.

It was the intent of this phase of the study to measure achievement in developing aesthetic sensitivity and preventing closure to new values in the Manhattanville Music Curriculum Program. The evidence shows that Manhattanville is somewhat successfully achieving its aims. The evidence also shows, albeit indirectly, that a test of aesthetic judgment can lend some weight to the evidence of this achievement.

The uses to which a test of aesthetic judgment in music may be put will vary with the needs of the learning situation. The primary need for evaluation in aesthetic education is to provide self knowledge, the knowledge as to what it is in an art work which affects the individual. Evaluation also permits the individual to position himself in relationship to his culture, and to control this cultural deprivation towards his own betterment.

Those teachers who find a need for an instrument to <u>predict</u> achievement will find a test of aesthetic judgments in music as serviceable as other types of tests for this purpose. It is true that all "talent" tests are, in a sense, achievement tests. Talent may be defined as the ability to adjust to the demands of the medium. The essence, here, is the rate of achievement or adjustment. Commercial tests, such as Gordon's <u>Musical Aptitude Profile</u>, which purport to measure ability rather than school achievement, probably have selected items which show the least effect of school music education. If formalized education does not account for individual differences, assumably some innate characteristic must. The value of the tests is that they have been validated by correlations with musicianly behavior. If it is a mark of a musician to recognize and appraise a well turned phrase the ability to do so by the young would set him aside from his less able classmates. If he is able to respond as a

musician at age ten while others typically reach the age of fifteen before attaining this ability, such evidence would suggest a precociousness worthy of special cultivation.

The primary use to which a test of aesthetic judgments in music should be put is in the assessment of curricular practices. As music educators turn their attention from the performance-dominated curriculum to one of conceptual development leading to increased aesthetic sensitivity, expectancies may well be catalogued as changes in the ability to comprehend musical ideas. Making aesthetic judgments is one behavioral manifestation of this ability.

Summary

From the evidence gathered in this research, several observations may be made:

1. The ability to make appropriate aesthetic judgments in music is highly correlated with teachers' estimates of the subjects' musicality.

2. The ability to make aesthetic judgments goes beyond perception and necessitates a response on the part of the listener in terms of his past musical experience.

3. Success on the California Test of Aesthetic Judgments in Music represents an adjustment to the demands of the medium.

4. Musical children are very much alike in their ability to make judgments of the appropriate usage of harmony, melody, rhythm, timbre, and musical style, regardless of the experimental program from which they are drawn.

5. The task of making aesthetic judgments in music is hierarchical in nature, and consequently a test of aesthetic judgments in music is not an adequate measure of musical achievements in grades lower than grade four.

From these observations some help may be offered to those in curriculum development who propose to make music education a form of aesthetic education. Music education would become aesthetic education when the emphasis is placed upon the behaviors which comprise the aesthetic experience. First and foremost of these behaviors is that of perceiving. Without perception

and its attendant behaviors of analyzing, conceptualizing, and producing, there is little chance that the aesthetic experience can exist.

Perception is not enough, for one must respond to music in order to experience it. Reacting to music includes the behaviors of evaluating and valuing. The combination of two aesthetic behaviors, perceiving and reacting, make possible the aesthetic experience, and all musical endeavor should be directed towards the development of these behaviors. Enlightened connoisseurship, the realization of value in music, as a component of the "good life" is nothing more nor less than this wedding of the cognitive and affective domains.

BIBLIOGRAPHY

1. Anastasi, A. Psychological Testing. New York: Macmillan, 1964.

2. Bugg, E. "Musical Discrimination in Musically Trained and Untrained Subjects," Journal of Educational Psychology, Vol. 35, pp. 498-499.

3. Farnsworth, Paul R. "Musical Taste, Its Measurement and Cultural Nature," Educational Psychology, II, No. 1. Stanford: Stanford University Press, 1952.

4. Farnsworth. "Are Musical Capacity Tests More Important than Intelligence Tests in the Prediction of the Several Types of Musical Grades?," Journal of Applied Psychology, XIX (1935), p. 347.

5. Gordon, Edwin. The Gordon Musical Aptitude Profile. Boston: Houghton Mifflin Co., 1965.

6. Hevner-Landsbury Oregon Music Discrimination Tests, University of Oregon Publications, Vol. 4, No. 6 (1963), pp. 115-130.

7. Holmes, J. A. "Reliabilities of Holmes Revision of Kwalwasser-Dykema Tests for Aptitude in Eight of the Elements of Music," Journal of Genetic Psychology, September, 1954.

8. Kwalwasser, J. and Dykema, P. W. Kwalwasser-Dykema Tests. New York: Carl Fischer, Inc., 1930.

9. Kyme, George H. "The Value of Aesthetic Judgments in the Assessment of Musical Capacity." Unpublished dissertation, University of California, Berkeley, 1954.

10. Kyme, George H. Developing Musicality in the Junior High School and the Contribution of Musical Composition to this Development. University of California, Berkeley, 1967 (H. E. W. 254).

11. Mainwaring, J. "An Examination of the Value of the Empirical Approach to Aesthetics," British Journal of Psychology, XXXII (1942).

12. M.E.N.C. "Tanglewood Symposium," M.E.N.C. Journal, Nov. 1967, p. 57.

13. Olsen, Roger. "A Comparative Study of the Effectiveness of Audio versus Audio-Visual Materials in Teaching Music in the Junior High School." Unpublished master's thesis, University of California, 1966. 96 pp.

14. Parker, Olin G. "A Study of the Relationship of Aesthetic Sensitivity to Musical Ability, Intelligence and Socio-economic Status." Unpublished doctoral dissertation, University of Kansas, 1961. 168 pp.

15. Palisca, Calude V. Seminar on Music Education. Cooperative Research Project No. G-013, Yale University, New Haven, Connecticut, 1963.

16. Rubin, Louis. "The Effects of Musical Experience on Musical Discriminations and Musical Preferences." Unpublished doctoral dissertation, University of California, 1952.

17. Taylor, C.H. "The Construction and Validation of Certain Experimental Measures of Musical Potentiality." Unpublished thesis, Teachers College, University of Cincinnati, 1940-43.

18. Thurstone, L.L. "The Calibration of Test Items," American Psychologist, II, No. 3 (1947), pp. 103-104.

19. Tryon, R.C. "Theory of the 'B.C. Try System': Statistical Theory." Unpublished paper. University of California, Berkeley, 1964.

20. Tryon, R.C. "Unrestricted Cluster and Factor Analysis, With Applications to the MMPI and Holzinger-Harmon Problems," Multivariate Behavioral Research, I, No. 2 (1966), pp. 229-244.

21. Tryon, R.C. and Bailey, D.E. "User's Manual of the 'B.C. Try System' of Cluster and Factor Analysis." Unpublished paper, University of California, Berkeley, 1965.

22. Tryon, R.C. and Bailey, D.E. "The 'B.C. Try' Computer System of Cluster and Factor Analysis," Multivariate Behavioral Research, I, No. 1 (1966), pp. 95-111.

23. Wing, Herbert. "Test of Musical Ability and Appreciation," British Journal of Psychology, Monograph Supplements. Cambridge University Press, XXXVII (1948).

APPENDIX

Description of the Test of Aesthetic Judgments in Music

Tape Transcription

This is a test to discover what you hear in music which causes you to enjoy it. You will hear two short pieces which are very much alike. While are are listening, you are to decide which one of them you like better. Sometimes the two pieces are exactly alike except for the rhythm. Sometimes it is the harmony that is different; sometimes it is the melody; and sometimes it is the way the music is played that is different.

You are to listen carefully and mark on your paper which of the two performances you prefer: A or B. If the two pieces are exactly alike, or if you cannot tell which is the better, do not guess but put a mark in the third column under C.

Let us listen to the first pieces together. Find the number on your answer sheet. Listen carefully.

1A
1B

Do you agree that the first performance is the better? Most people do, so put a mark through the A opposite 1.

Now listen to the next example:

2A
2B

Here the better performance is the second, so you should make a mark through B opposite 2.

Now listen to number 3:

3A
3B

Did you prefer A or B? The answer is C, for the two performances were exactly alike, so put a mark through C opposite 3.

Remember: do not guess. If the two pieces are exactly alike or if you cannot tell which is the better, put a mark under C.

Now, if you understand how to mark your paper, we shall go on with the test.

Items Used for Illustration

Schuman: Kinderscenen, Op. 15, No. 2. Rhythm; A.

This is an obvious rhythmic distortion; the notes of the original melody were kept, but we employed an entirely different rhythm to accompany it. We began with a triplet figure, while the original used a quarter note. Also, we used fewer dotted rhythms. In general, our purpose in the distortion was to remove the emphasis from the places where it properly belonged. In the second measure we interpolated a waltz accompaniment.

Schumann: Kinderscenen, Op. 15, No. 13. Harmony; B.

Harsh dissonance employed at random, without any regard for the rules of chord structure, is employed in this harmonic change. Notes are used which have no relation to the chord into which they are inserted. In addition, we break a rule ofharmony by using a stationary bass over the measure line.

Joplin: Maple Leaf Rag, second section. Rhythm; A.

The mutilation retains the left-hand two-step alternating bass, but all syncopations and cycles-of-3 are removed from the right hand, substituting square, on-the-beat melody notes with straight quarters and eighths.

Items in the Test and Key for Scoring

1. Haydn: Sonatina. Melody; A.

In this composition the melody has been distorted in the first, third, fifth, sixth, and seventh measures. The changes are obvious, inasmuch as they are contrary to the harmonic implications in the bass. In the next to the last measure we have two major seventh skips which are particularly unmelodic, and in the fifth measure we have an "e flat" melody note against a predominating C major harmony.

2. Schumann: Scherzino, Op. 26, No. 3. Rhythm; A.

The change here is in the rhythm. The theme of the original is lively, and the rhythm very decided. In the mutilation, the dotted rhythm is removed, except in the last measure where its presence is incongruous, and a lifeless and awkward rhythm is substituted.

3. Liadoff: Prelude in C. Major, Op. 40, No. 1. Harmony; B.

The harmony in the bass of this composition is modified by replacing the existing chord structure with harsh dissonances out of character with the subdued, soporific suggestiveness of the original; substituting "a flat" and "f sharp" for a C major triad is an example of the type of change.

4. Dreyshock: Gavotte. Melody; A.

One of the obvious features of change in this modification is the distance which the melody skips in going from one note to another.

5. Schumann: Little March. Rhythm; B.

In the mutilated version the accent is changed so that the music appears to start with an anacrusis.

6. Haydn: Andante Grazioso. Harmony; A.

The mutilation is in this case relatively unimportant until the next to the last measure. Previous to this, there is a slight harmonic change in the second measure, and in the fifth, a melodic change from "e flat" to "e sharp." However, in the seventh measure, we make two startling digressions from the original: first, we change a "d" minor root position chord to an "f" major 6/4 chord; second, we substitute for a tonic 6/4, a dominant chord of "f." These substitutions, although not far removed from the rest of the composition in key relationship, seem unwarranted because they are not inserted in places where they sound well in relation to the preceding and following chords.

7. Schumann: Happy Farmer. Rhythm; A (Tempo).

The piece is played normally at 96 beats per minute, but the mutilated version was played at 136 beats per minute.

8. Marpurg: La Voltigeuse. Rhythm; A.

We take the joyousness, the sprightliness, out of this composition by changing the rhythm. The original is in 2/4 tempo, with many sixteenth notes which are all to be lightly or quickly played. We substitute eighth and quarter notes for these sixteenth notes, in this way dragging out the piece until the virility of the original is entirely gone.

9. Chopin: Mazurka, Op. 7, No. 5. Rhythm; B.

This is one of the most obvious mutilations in the test. We remove all trace of the rhythmic variation in the original by having in the melody a continuous succession of 36 quarter notes. We use in the bass all sorts of unrelated accompaniments to support this tedious melody, including a waltz rhythm, an eighth note single accompaniment, and then an eighth note chordal structure. We follow this variation by a return to

the mazurka character, and finally close with an arpeggio figure and some simple chords.

10. Beethoven: Allegretto, from Sonata, Op. 27, No. 2. Harmony; A.

Beethoven has here a bass line which is excellently constructed; it is interesting in itself and has a complete unity with the rest of the composition. We distort this bass both by unmelodic skips and by changing the harmonic implications; as an example of the former, in going from the fifth to the sixth measures, we progress from "a flat" to "d flat." In other ways we spoil the harmony: by consecutive fifths in the second and third measures, and by consecutive octaves in the sixth and seventh. In general, all our harmonic progressions are without distinction as compared with those of the original.

11. Schubert: Ave Maria. Style; A.

In the mutilation of this violin rendition the vibrato was judged too rapid--causing the music to sound excited and sharp in pitch. Moreover, the melody was contracted to produce a 7-1/2 measure phrase.

12. Debussey: Le Fille aus Cheveux de Lin. Style; B.

In this item the phrasing (nuances) of two professional violinists was contrasted.

13. Bach: Double Concerto in D. Minor. Style; B.

This item was one of balance between soloists and string orchestra accompaniment. In the first presentation the solo voices are played on the harpsichord. In the second performance, the customary two violins perform.

14. Folk song: I Wonder as I Wander: Style; A.

In contrast to the simplicity of a young soprano singing this melody, an overly theatrical folk singer sang this unassuming Appalachian tune.

15. Cello Improvisation. (From the Manhattanville Music Curriculum Program). Melody; A.

The first performance was more coherent by customary standards. Leaps were followed by a change of melodic direction and expected voice leadings were observed.

16. Brahms: Symphony I (4th Movement, First Theme). Form; B.

Here, by splicing the tape, the consequent phrase was placed before the antecedent phrase in the mutilation.

17. Debussey: Little Shepherdess. Intonation; B.

In the less preferable version of this flute solo, the flautist pulled the tuning joint one inch. The piano at the same time played the accompaniment an octave lower than written.

18. Manhattanville Improvisation. Form; A.

This instrumental improvisation contrasted two forms of an exercise in tone colors. The development of the idea was the chief factor contributing to the preference of the first performance.

19. Bach: Aria (from Suite in D for Strings). Style; B.

The string orchestra version of this aria was judged more appropriate than a choral group humming it, primarily due to the predominance of a snare drum accompaniment in the Swingle Singers arrangement.

20. Beethoven: Symphony No. V. Form; B.

Some of Beethoven's rough sketches from his notebook were put together for the less desirable form of these initial 32 measures. Moreover, the presence of a flute in the opening measures made this arrangement ridiculous compared to the final version.

21. Mozart: Sonata IV in B Flat Major. Melody; A.

This lovely melody of Mozart is spoiled through the insertion of notes which do not carry out the set of the previous structure. Mozart ascends to "g" in the second measure; we remain on "e flat." He repeats his theme at the beginning of the fourth measure; we substitute the melody note "b flat" for the original "f." In other words, the normal expectation is distorted until the result becomes entirely lacking in form.

22. Raff: Rigaudon, Op. 204. Rhythm; B.

This composition digresses from the original rhythm in obvious ways. Two eighth notes become half notes in the spoiled version; in the same manner quarter notes change to eighths, regular rhythms become dotted, grace notes are eliminated, until scarcely a vestige of the sprightly original is left. The accompaniment is made unwieldy by inserting into the fast moving quarter notes of the original occasional triplets, which slow up the vigorous movement of the Rigaudon.

23. Grieg: Sarabande, from the Suite Aus Halberg's Zeit. Harmony; A.

The harmony in this item is modified through the insertion of unrelated dissonance. We insert chords at random which have no tonal relation with what precedes or follows. On the other hand, in the sixth and seventh measures we have removed Grieg's interesting progressions so that these measures seem particularly incongruous in comparison with the earlier ones.

24. Scriabine: Prelude in B Major. Melody; A.

This spoiled melody is made to sound peculiar by making several of the notes clash with the established harmonic structure. The changes made are so pronounced that the melodic line played by itself sounds odd, unusual, as if the notes were chosen simply at random without any thought of a harmonic background for them.

25. Solovyeff: In the Fields. Rhythm; B.

The outstanding rhythmic characteristic of this composition, which is in 6/8 time, is a syncopation on the second and fifth beats of each measure. In the mutilation we remove this feature entirely. In the first place, we substitute for the 6/8 rhythm a combination of 3/4 and 4/4 time. We accompany this changed melody in varied ways: first, by simply using half notes, then by a waltz rhythm, again by triplets, and so on with similar variations throughout the whole piece.

26. Scott: Serenata, Op. 67, No. 2. Harmony; A.

We spoil the subtle harmonic procedure by using even more dissonance than the original, harsh sounds which are unorganized, which have no place in an ordered system. For example, we will use the "d flat" triad as basic harmony, and then in the melody will continue the original melodic fragment, "e," "f sharp," and "g sharp." The result is a conglomeration which has no past association for us, and therefore seems strange and peculiar.

27. Beethoven. Sonata, Op. 79. Melody; A.

At two different times in this composition we restrict the range of the melody. In the second and third measures the spoiled version has a limit of two notes while the original uses a range of five; similarly, in the fifth and sixth bars the original has a compass of seven notes, while the mutilation is restricted to two. As a result, the modification is without melodic variety or contrast.

28. Mozart: Die Entfuhring, aus dem Serail. Melody; A.

This is a change in which the melody is made uninteresting by keeping it within a range of five notes instead of using the compass of an octave as does the original. For example, in the third measure the correct version ascends gradually to a high "g" while the spoiled version keeps repeating the notes "b" and "c." This obviously makes the tune dull and pointless. In the sixth and seventh measures, through similar repetition, this same effect is achieved again.

29. A. Koretschenkso. Au Mois de Mai, Op. 22, No. 2. Harmony; B.

Here the variation between the original and the spoiled version is in the harmony. We keep the first measures mostly in the tonic key instead of having them progress to the dominant as in the original; however, we get a strange effect by introducing mediant harmony on the last beats of the first and second measures. The third measure has unrelated dissonance inserted into it, while the final measures are made dull through constant repetition of dominant harmony.

30. Buchner: Cheerfulness, Op. 12, No. 1. Harmony; A.

In this composition the movable bass line provides a sympathetic background for the bright character of the original melody. It is the harmony of this bass which we mutilate by removing the unity and flow of its movement. The spoiled version repeats over and over again the same notes; it does not progress continuously to any given point, as in the original.

31. Mozart. Sonata II. Rhythm; A.

The sharp rhythmic contour of the original is dismantled in this mutilation through the elimination of its existing sharp contrasts. The first measure remains intact but the following modification includes an elimination of grace notes, thirty-second notes, and a substitution of eighth for sixteenth notes.

32. Mascagni. Wine Song, from Cavalleria Rusticana. Form; A (R or M).

The fifth measure alone is the object of mutilation in this excerpt from the Opera Cavalleria Rusticana. Where in the fifth measure Mascagni reaches the climax of his eight measure theme, we instead insert a series of four triplets, in this way supplanting his high point by a series of notes, not only unmelodic but unrelated to the previous structure.

33. Gershwin: Fascinatin' Rhythm. Harmony; A.

As a variant from the original, the harmony was changed in the sequential fourth, fifth, sixth, and seventh measures.

34. Kyme: Minuet. Rhythm; A.

This minuet was played in good minuet style and then with misplaced accents which turned the gracefulness of the dance form into a monotonous 4/4 march.

35. Kern: All the Things You Are. Melody; B.

This beautiful song was marred in the mutilated performance by lowering the third melodic and harmonic sequence one-half step.

36. Dawes: Melody in A. Harmony; B.

A tonic harmony was substituted throughout for the colorful harmony originally employed in this composition.

37. Duke and Harburg: April in Portugal. Rhythm; A.

As compared to the first performance of this rhythmically interesting piece, the accompaniment of the altered version was one of confused rhythms.

38. Kyme: Ostinato. Harmony; B.

This piece with its constantly repeated bass pattern was performed so that the beauty due to the consistency of the ostinato was sacrificed in the repeated version for another bass pattern that alternated between the melodic and pure forms of the minor scale.

39. Alter: Manhattan Serenade. Harmony; A.

As an inferior version, the melody was altered. The octave basso repetition of the third measure was played a major second lower than originally intended.

40. Gershwin: Rhapsody in Blue. Rhythm; A.

In the theme from this composition, the chromatic rhythmical obbligato in the third through sixth measures was changed to a six-eight rhythm, thus emasculating this interesting figure.

41. Gershwin: American in Paris. Rhythm; A.

The melody was played legato and its rhythm altered in the inferior performance of this item.

42. Sweet: Fight On. Harmony; B.

This item comprised two performances of this college football song. In the second performance the bass notes were altered by flatting the seventh tone of the scale.

43. Guilty. Melody; A.

In the least preferred version, an extra measure was interpolated at the fourth measure which extended the sequence into the wrong key.

44. I Believe. Form; B.

The repetitious first two measures of I Believe was extended another half measure, thus carrying the repeated one note to a point of nausea. This was compared to the original edition.

45. Coca Cola. Harmony; B.

The theme song Coca Cola was played with two harmonies, the first of which the theoretician would most likely prefer.

46. Hagen: Harlem Nocturne. Melody; B.

Major thirds were substituted for the expected minor thirds in the melody of this piece, thus presenting a problem of consistency of mode between the melody and its harmony.

47. Hindemith: Piano Sonata No. 2, First Movement, last page. Harmony; A.

In the mutilation, rhythms are left as is, but Hindemith's atonal harmonies with their quick color shifts are replaced by an innocuous version which sticks closely to the diatonic harmonies of C minor and E-flat major.

48. Milhaud: Saudades do Brazil, No. 7: "Corcovado," opening. Harmony; B.

In the mutilation, the tango (habanera) rhythm has been left intact, but the melody, in D major, has been harmonized in that key throughout, instead of in G major, as Milhaud wrote it. At the place where Milhaud shifts to A-flat major (with appropriate bi-chordal harmonies above), a version sticking to D major is provided. The purpose of the mutilation is to substitute straight diatonic harmonies for the bi-chordal and bi-tonal original.

49. Schoenberg: Six Little Piano Pieces, Op. 19, No. 4 (entire). Melody; B.

The mutilation retains the rhythmic scheme in general, but uses diatonic F major instead of Schoenberg's atonal procedures. This example stands midway between Schoenberg's early romanticism and his later 12-tone style.

50. Ravel: Sonatine, Movement No. 1, Secondary Theme of Exposition. Harmony; B.

The mutilation treats Ravel's exact theme with two chords only--tonic and dominant in E major--whereas Ravel's harmonization is parallel shifting major triads, with no particular key, and a strong modal touch. Needless to say, consecutive 5ths are featured in both right and left hand.